

Edited by

Hirak Ranjan Dash, Kelly M. Elkins and Noora Rashid Al-Snan



Next Generation Sequencing (NGS) Technology in DNA Analysis



Next Generation Sequencing

(NGS) TECHNOLOGY IN DNA ANALYSIS

This page intentionally left blank

Next Generation Sequencing (NGS) TECHNOLOGY IN DNA ANALYSIS

Edited by

HIRAK RANJAN DASH

Department of Forensic Science, National Forensic Sciences University, Delhi, India

KELLY M. ELKINS

Forensic Science Program, Department of Chemistry, Towson University, Towson, MD, United States

NOORA RASHID AL-SNAN

Forensic Science Laboratory, Directorate of Forensic Science, General Directorate of Criminal Investigation and Forensic Science, Ministry of Interior, Manama, Bahrain





ACADEMIC PRESS

An imprint of Elsevier

Academic Press is an imprint of Elsevier
125 London Wall, London EC2Y 5AS, United Kingdom
525 B Street, Suite 1650, San Diego, CA 92101, United States
50 Hampshire Street, 5th Floor, Cambridge, MA 02139, United States

Copyright © 2024 Elsevier Inc. All rights reserved, including those for text and data mining, AI training, and similar technologies.

No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher. Details on how to seek permission, further information about the Publisher's permissions policies and our arrangements with organizations such as the Copyright Clearance Center and the Copyright Licensing Agency, can be found at our website: www.elsevier.com/permissions.

This book and the individual contributions contained in it are protected under copyright by the Publisher (other than as may be noted herein).

Notices

Knowledge and best practice in this field are constantly changing. As new research and experience broaden our understanding, changes in research methods, professional practices, or medical treatment may become necessary.

Practitioners and researchers must always rely on their own experience and knowledge in evaluating and using any information, methods, compounds, or experiments described herein. In using such information or methods they should be mindful of their own safety and the safety of others, including parties for whom they have a professional responsibility.

To the fullest extent of the law, neither the Publisher nor the authors, contributors, or editors, assume any liability for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions, or ideas contained in the material herein.

ISBN: 978-0-323-99144-5

For information on all Academic Press publications visit our website at https://www.elsevier.com/books-and-journals

Publisher: Stacy Masucci
Acquisitions Editor: Kattie Washington
Editorial Project Manager: Susan Ikeda
Production Project Manager: Neena S. Maheen
Cover Designer: Greg Harris

Typeset by TNQ Technologies



Dedications

This book is dedicated to my beloved parents and my precious kids (Moni and Khalifa) who always have faith in me

Noora Rashid Al-Snan

This book is dedicated to my parents and my husband Tim, with gratitude for inspiring and encouraging me to innovate and test new technologies

Kelly M. Elkins

I would like to dedicate this book to my mother Manorama Dash and my father Janmejaya Dash

Hirak Ranjan Dash

This page intentionally left blank

Contents

Contributors *xvii*
Preface *xxi*
Acknowledgments *xxiii*

SECTION 1 Useful applications of NGS-based forensic DNA analysis

1. Transition of capillary electrophoresis to next generation sequencing for forensic DNA analysis: Need of the hour **3**
Noora Rashid Al-Snan

Introduction 3
DNA sequencing platforms 5
Overview of the methodology 6
Template preparation 6
Sequencing and imaging 6
Data analysis 6
The impact of NGS on genomics research 7
Genomic analysis 8
Targeted genomic resequencing 8
Metagenomics 9
The impact of NGS on forensic genetics 9
Short tandem repeats sequencing 10
The initial commercial NGS kits for forensic genetics 11
Current status of next generation sequencing 12
Recommendations 15
Concluding remarks 16
References 16

2. Using conventional STR technology in analyzing biological transfer evidence **21**
J. Thomas McClintock

Introduction 21
Touch DNA (or trace DNA) 22
Direct and indirect transfer of biological evidence 23
Effects on transfer and shedder status (assessment of shedder status and implications for secondary DNA transfer) 26

Touch DNA and wrongful convictions 31
A new awareness 32
References 33
Further reading 34

3. Overview of NGS platforms and technological advancements for forensic applications 35
Anna Barbaro

Introduction 35
Next-generation sequencing platforms 36
NGS forensic applications 40
Forensic NGS commercial systems 48
NGS advantages and challenges 50
References 52

4. Processing of biological samples for forensic NGS analysis 59
Sulekha Coticone and Camila Garcia

Overview of next-generation sequencing 59
NGS versus Sanger sequencing: What's the difference? 59
Next-generation sequencing in forensic science 59
Overview of the sample preparation process for next-generation sequencing 60
Sample preparation methods 61
DNA extraction 62
DNA quantitation 64
Library preparation 65
Library normalization 68
References 69

5. Commercial kits commonly used for NGS based forensic DNA analysis 73
Tugba Unsal Sapan

Introduction 73
Development of next-generation sequencing systems 73
Use of NGS in forensic sciences 76
Conclusion 81
References 81

6. Applications of nanopore sequencing for forensic analysis 85
Olivier Tytgat and Filip Van Nieuwerburgh

Introduction 85
Nanopore sequencing 86
Forensic applications of nanopore sequencing 88

Opportunities 93
Challenges 94
Conclusions 95
References 95

7. Microhaplotypes analysis for human identification using next-generation sequencing (NGS) 99
Mohammed H. Albujja

Introduction to forensic DNA analysis 99
DNA typing techniques used in forensic DNA analysis 100
DNA sequencing 101
Microhaplotypes as a tool for human identification 102
Next-generation sequencing for microhaplotypes analysis 106
Validation of NGS-based microhaplotypes analysis 108
Future directions and opportunities in NGS-based microhaplotypes analysis 110
Ethical considerations for NGS-based microhaplotypes analysis 113
Potential applications of NGS-based microhaplotypes analysis 114
Future directions of NGS-based microhaplotypes analysis 116
Conclusion 117
References 118

8. Tools and techniques of using NGS platforms in forensic population genetic studies 121
Bofeng Zhu and Chunmei Shen

Overview of the concept of population study 121
Forensic parameters of population genetics 123
Commonly used markers of population study on NGS platform 133
Statistical analytical methods of population genetic study 151
Learning practices of population study based on NGS platform for developing a 40 ancestry informative SNPs (AI-SNPs) multiplex panel for forensic ancestral inference on 17 different continental populations 157
Summary 165
Acknowledgments 165
References 165

SECTION 2 Useful applications of NGS-based forensic DNA analysis

9. Applications of NGS in analysis of autosomal STRs 175
Hannah Berry

Introduction 175
Next-generation sequencing 175

Application of NGS in autosomal STR analysis 176
Comparing loci amplified using different kits for NGS applications, including two kits for CE applications 178
Population studies of autosomal STRs using NGS 181
Conclusion 183
References 183

10. Differentiating monozygotic twins using NGS 185
Cassandra Skrant, Ilya Grabylnikov, Kelly M. Elkins and Cynthia B. Zeller

Introduction to DNA analysis methods 185
Mitochondrial genome sequencing 185
Nuclear genome sequencing 187
RNA profiling: microRNAs 188
High-resolution melting to determine DNA methylation 188
Detecting DNA methylation with pyrosequencing 189
Genome-wide DNA methylation profiling 190
Conclusion 192
Acknowledgments 193
References 193

11. Application of NGS in maternal genome analysis in ancient human remains 195
Muhammad Irfan, Khushbukhat Khan and Naila Khan

Introduction 195
What stories ancient DNA analysis tell us? 196
Challenges faced in studying ancient DNA 198
Next-generation sequencing for maternal genome analysis in ancient samples 199
Next-generation sequencing based on ancient sample mtDNA in evolutionary studies 204
Recent developments 208
Future perspective 209
References 210

12. Application of NGS technology for parentage testing and relatedness analysis 219
Andrei Semikhodskii

Introduction 219
Parentage testing 227
Kinship testing 234

Conclusions 235
References 236

13. Forensic relevance of SNP analysis in next-generation sequencing 243
Seema Malhotra and Shivani Sehgal

Introduction 243
Characterization and origin of SNPs 244
Conventional SNP genotyping assays 245
Benefits of next generation sequencing technology over conventional genotyping assays 245
Forensic application of SNP markers using NGS technology 258
Limitations of SNP analysis in forensics 267
Conclusions 268
References 269

14. Forensic analysis of hair using NGS technology 279
Ruth M. Benitez and Kelly M. Elkins

Introduction to hair analysis in forensic science 279
Mitochondrial DNA analysis 280
Nuclear DNA analysis 281
Application of NGS for hair analysis 283
Conclusion 286
References 287

15. Forensic DNA phenotyping using next-generation sequencing 289
Mahdi Haidar, Fatema Mousawi and Abdullah K. Al-Matrouk

Introduction 289
Progression to next-generation sequencing 291
Phenotypic-informative SNP in forensics 292
Externally visible characteristics 293
Next-generation sequencing in forensics 300
References 303

16. Forensic DNA phenotyping in the next-generation sequencing era 311
Thássia Mayra Telles Carratto, Maria Luiza Guimarães de Oliveira and Celso Teixeira Mendes-Junior

Introduction 311
Forensic DNA Phenotyping 312

Final remarks 326
References 327

17. Forensic applications of epigenetic (DNA methylation) markers through NGS 337

Meenu Ghai

Introduction 337
Emerging MPS methods for forensics 338
Body fluid identification by MPS 340
Age estimation by MPS 344
Improvements in the application of epigenetic markers via MPS 350
Future considerations 350
References 351

18. Forensic applications of NGS-based microRNA analysis 357

Noora Rashid Al-Snan

Introduction 357
miRNA and body fluid identification 358
Discrimination of monozygotic twins 359
miRNAs and time since death determination 362
Wound age estimation using miRNA 364
Organ tissue identification using miRNA 366
miRNA and drowning 367
miRNA and sepsis 368
Other potential future application 368
Potential limitations 369
Concluding remarks 369
Acknowledgments 369
References 370

19. NGS profiling of water and soil microbial DNA in forensic science 375

Jillian E. Malbrough and Kelly M. Elkins

NGS approaches for microbial profiling of soil and water 375
Extraction methods for obtaining microbial DNA from soil and water 377
Quantitation methods for microbial DNA 381
NGS methods for microbial DNA analysis 381
Application of microbial DNA analysis in geolocation in the United States 382
Ongoing challenges and conclusion 383
References 383

20. Applications of NGS in analysis of challenging samples 387
Safia A. Messaoudi

Introduction 387
Application of next-generation sequencing technology in forensic science 388
Panels developed for the forensic samples using the NGS technologies 388
DNA damage and degradation 392
Challenging DNA evidence 394
Previous types of technologies used to analyze challenging DNA 395
Application of NGS in challenging DNA evidence 398
Conclusion: The main challenges of NGS in forensic applications 403
References 404

21. Detection of human body fluid through mRNA analysis using NGS 413
Quentin Gauthier

Introduction 413
Body fluid identification 414
Sequencing approach 417
Beyond just body fluids 418
Analysis methods for mRNA data 419
Future of mRNA body fluid identification 420
References 421

22. NGS-based detection and differentiation of forensically relevant body fluids using conventional, molecular, and microbial techniques 425
Hirak Ranjan Dash

Background 425
Existing technologies for human body fluid detection 427
Conclusions 444
List of abbreviations 444
References 445

23. Application of next-generation sequencing technology and investigative genetic genealogy to solving cold cases 451
Daria Smirnova and Kelly M. Elkins

Introduction 451
Methods 452
Results 459
Discussion and conclusion 462

References 464
Case Reports 465

SECTION 3 Challenges in implementing NGS in forensic laboratories

24. Troubleshooting and challenges of Next-generation sequencing technology in forensic use 471
Eugenio D'Orio, Aurora Toso, Paola Montagna and Maria Denise Gemmellaro

Forensic genetics: timeline of technology evolution 471
Current limitations and problem solving in forensic genetics 473
NGS: from diagnostics to forensic applications 475
NGS in forensics and the issue of privacy: a global overview 479
Applications of NGS technology in forensic entomology 481
References 483

25. Validation of NGS for casework at forensic DNA laboratories 485
Deborah Soares Bispo Santos Silva

Introduction 485
NGS kits and platforms for generating DNA profiles 487
Milestones for NGS in forensic casework 488
Validation of NGS-based DNA analysis methods 490
Funding for forensic laboratories to adopt and validate NGS technology 505
References 506

26. Challenges in using genetic genealogy in forensics 511
Praveen Kumar Yadav and Sweety Sharma

Introduction 511
How forensic DNA profiling differs from forensic genetic genealogy 511
How investigative genetic genealogy works 512
Challenges in using genetic genealogy 515
Way forward 526
Conclusions 527
References 527

27. The role of artificial intelligence and machine learning in NGS 531
Khaled Ouanes

What is next-generation sequencing? 531
Next-generation sequencing for forensics 531

What is artificial intelligence? Why is it useful for forensics? 533
What is machine learning? 534
How do artificial intelligence and machine learning help with NGS? 535
Challenges of utilizing AI and ML in NGS 536
Conclusion 536
References 537
Further reading 538

28. Ethical issues of forensic application of NGS technology 539
Theresa M. DeAngelo and Kelly M. Elkins

Introduction to forensic DNA analysis 539
Ethics and forensic DNA typing 540
How does forensic genetic genealogy differ from familial DNA searches? 541
Misuse of tools indicated for forensic use? 542
Personal genomics companies and consumer data in the forensic setting 543
Forensic NGS and genetic genealogy 544
Next-generation sequencing and environmental impact 545
Next-generation sequencing education needs 546
Conclusion 548
References 548

29. Ethical, legal, and social challenges of next-generation sequencing technologies (NGS) in forensic criminal identification 551
Helena Machado, Laura Neiva, Rafaela Granja and Susana Silva

Introduction 551
Gaps in legislation and regulation 553
Access to databases that hold full-genome data 554
Data protection 556
Algorithms, privacy, and discrimination 557
Societal trends framing NGS technologies 559
Conclusion 563
Acknowledgment 564
References 564

Index *571*

This page intentionally left blank

Contributors

Abdullah K. Al-Matrouk
Forensic Laboratory Department, General Department of Criminal Evidence, Ministry of Interior, Kuwait City, Kuwait

Noora Rashid Al-Snan
Forensic Science Laboratory, Directorate of Forensic Science, General Directorate of Criminal Investigation and Forensic Science, Ministry of Interior, Manama, Bahrain

Mohammed H. Albujja
Department of Forensic Sciences, Naif Arab University for Security Sciences, Riyadh, Saudi Arabia

Anna Barbaro
Department of Forensic Genetics, Studio Indagini Mediche E Forensi (SIMEF), Reggio Calabria, Italy; Worldwide Association of Women Forensic Experts (WAWFE), Rome, Italy

Ruth M. Benitez
Department of Chemistry, Forensic Science Program, Towson University, Towson, MD, United States

Hannah Berry
Towson University, Towson, MD, United States

Deborah Soares Bispo Santos Silva
Department of Chemistry, Hofstra University, Hempstead, NY, United States

Thássia Mayra Telles Carratto
Departamento de Química, Laboratório de Pesquisas Forenses e Genômicas, Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, São Paulo, Brazil

Sulekha Coticone
Department of Chemistry and Physics, Florida Gulf Coast University, Ft Myers, FL, United States

Eugenio D'Orio
Bio Forensics Research Center, Ischia, Italy

Hirak Ranjan Dash
Department of Forensic Science, National Forensic Sciences University, New Delhi, India

Theresa M. DeAngelo
Maryland State Police Forensic Sciences Division, Pikesville, MD, United States

Kelly M. Elkins
Department of Chemistry, Forensic Science Program, Towson University, Towson, MD, United States

Camila Garcia
Department of Chemistry and Physics, Florida Gulf Coast University, Ft Myers, FL, United States

Quentin Gauthier
Alexandria, VA, United States

Maria Denise Gemmellaro
Department of Biology, Kean University, Union, NJ, United States

Meenu Ghai
Department of Genetics, School of Life Sciences, University of KwaZulu Natal, Durban, KwaZulu-Natal, South Africa

Ilya Grabylnikov
Department of Chemistry, Forensic Science Program, Towson University, Towson, MD, United States

Rafaela Granja
Communication and Society Research Centre, University of Minho, Braga, Portugal

Maria Luiza Guimarães de Oliveira
Departamento de Genética, Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, São Paulo, Brazil

Mahdi Haidar
Kuwait Identification DNA Laboratory (KIDL), General Department of Criminal Evidence, Ministry of Interior, Kuwait City, Kuwait

Muhammad Irfan
Department of Healthcare Biotechnology, National University of Science and Technology Islamabad, Atta-ur-Rahman School of Applied Biosciences, Islamabad, Pakistan

Khushbukhat Khan
Department of Healthcare Biotechnology, National University of Science and Technology Islamabad, Atta-ur-Rahman School of Applied Biosciences, Islamabad, Pakistan

Naila Khan
Department of Healthcare Biotechnology, National University of Science and Technology Islamabad, Atta-ur-Rahman School of Applied Biosciences, Islamabad, Pakistan

Helena Machado
Institute for Social Sciences, University of Minho, Braga, Portugal

Jillian E. Malbrough
Department of Chemistry, Forensic Science Program, Towson University, Towson, MD, United States

Seema Malhotra
School of Forensic Sciences, National Forensic Sciences University (NFSU), Delhi Campus, Delhi, India

J. Thomas McClintock
Liberty University, Department of Biology and Chemistry, Lynchburg, VA, United States

Celso Teixeira Mendes-Junior
Departamento de Química, Laboratório de Pesquisas Forenses e Genômicas, Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, São Paulo, Brazil

Safia A. Messaoudi
Department of Forensic Science, College of Criminal Justice, Naif Arab University for Security Sciences, Riyadh, Saudi Arabia

Paola Montagna
Bio Forensics Research Center, Ischia, Italy

Fatema Mousawi
Kuwait Identification DNA Laboratory (KIDL), General Department of Criminal Evidence, Ministry of Interior, Kuwait City, Kuwait

Laura Neiva
Institute for Social Sciences, University of Minho, Braga, Portugal; Communication and Society Research Centre, University of Minho, Braga, Portugal

Khaled Ouanes
Health Informatics Department, Health Sciences College, Saudi Electronic University, Dammam, Saudi Arabia

Shivani Sehgal
School of Forensic Sciences, National Forensic Sciences University (NFSU), Delhi Campus, Delhi, India

Andrei Semikhodskii
Medical Genomics Ltd, London, United Kingdom

Sweety Sharma
Department of Forensic Biology and Biotechnology, School of Forensic Sciences, LNJN National Institute of Criminology and Forensic Science, National Forensic Science University, New Delhi, India

Chunmei Shen
Department of Laboratory Medicine, Nanfang Hospital, Southern Medical University, Guangzhou, Guangdong, China

Susana Silva
Institute for Social Sciences, University of Minho, Braga, Portugal; Centre for Research in Anthropology, University of Minho, Braga, Portugal

Cassandra Skrant
Department of Chemistry, Forensic Science Program, Towson University, Towson, MD, United States

Daria Smirnova
Department of Chemistry, Forensic Science Program, Towson University, Towson, MD, United States

Aurora Toso
Bio Forensics Research Center, Ischia, Italy

Olivier Tytgat
Ghent University, Laboratory of Pharmaceutical Biotechnology, Ghent, Belgium

Tugba Unsal Sapan
Üsküdar University, Faculty of Engineering and Natural Science, Department of Forensic Sciences, Institute of Addiction and Forensic Sciences, Istanbul, Turkey

Filip Van Nieuwerburgh
Ghent University, Laboratory of Pharmaceutical Biotechnology, Ghent, Belgium

Praveen Kumar Yadav
Centre for Integrated Studies, Cochin University of Science and Technology, Kochi, Kerala, India

Cynthia B. Zeller
Department of Chemistry, Forensic Science Program, Towson University, Towson, MD, United States

Bofeng Zhu
Guangzhou Key Laboratory of Forensic Multi-Omics for Precision Identification, School of Forensic Medicine, Southern Medical University, Guangzhou, Guangdong, China

Preface

Writing a science book on recently advanced topics was not an easy journey, but it was worth all the time and challenges since we know that it will be used to serve justice to those who committed crimes and help to sustain peace with the power of science. This book was made in a continuous work over more than 2 years (since 2021), starting from writing a proposal, gathering the best authors, and following up with authors with deadlines. The hardest job was scheduling the perfect timing between different editors living in different sides of the world with extreme various time zones; when it was too early for one, it was too late for the other one.

Next Generation Sequencing (NGS) Technology in DNA Analysis explains and summarizes NGS technological applications in the field of forensic DNA analysis. This book covers the shift from capillary electrophoresis (CE) to NGS platforms and the fundamentals of NGS technologies, applications, and advancements. Sections provide an overview of NGS technology and forensic science, including information on processing biological samples for forensic analysis, sequence analysis and data analysis software for the analysis of NGS data. Next, the book explores the useful applications of NGS based forensic DNA analysis and covers the validations and interpretation guidelines of NGS workflows. Though most of the forensic science laboratories are in the process of migrating from CE based techniques to NGS based techniques, the limitations and future challenges of NGS technology in forensic science has also been explained in the book.

With chapter contributions from an international array of experts and the inclusion of practical case studies, this book is the most useful reference for academics and researchers in genetics, biotechnology, bioinformatics, biology, and medicine as well as forensic DNA scientists and practitioners who aim to learn, use, apply, and validate NGS-based technologies. We hope that this book will shed light to the scientists and academics that seek knowledge and offer them the answers they pursue. We look forward to conversations with our readers.

Happy Reading!!!!

Hirak Ranjan Dash
Kelly M. Elkins
Noora Rashid Al-Snan
May 2023

This page intentionally left blank

Acknowledgments

We would like to express our sincere gratitude to all the authors for their contributions in the book. Their time and efforts they provided throughout the years are what made our book a reality. Dr. Hirak Ranjan Dash would like to acknowledge his wife Banani and daughter Ditsa for their continual support and encouragement throughout the journey of writing and editing the book. Also we would to thank the Elsevier team who supported us with their help and assistance in all the stages by reducing all the challenges to a minimum level. It has been a great journey to all of us.

This page intentionally left blank

SECTION 1

Useful applications of NGS-based forensic DNA analysis

This page intentionally left blank

CHAPTER 1

Transition of capillary electrophoresis to next generation sequencing for forensic DNA analysis: Need of the hour

Noora Rashid Al-Snan
Forensic Science Laboratory, Directorate of Forensic Science, General Directorate of Criminal Investigation and Forensic Science, Ministry of Interior, Manama, Bahrain

Introduction

It has been more than a decade since the first next generation sequencing (NGS) technology appeared, dramatically altering the way genetic research is conducted and ushering in a new era of high-throughput genomic analysis. The terms NGS, massively parallel sequencing (MPS), and deep sequencing are all used to describe a DNA sequencing method that has revolutionized genomic research. The full human genome can be sequenced in a single day using NGS (Behjati & Tarpey, 2013). Today, full genomes are mapped and published almost weekly, at an increasing rate and at a lower cost (Børsting & Morling, 2015). NGS methods and platforms have matured over the last decade, and the quality of the sequences has reached the point where NGS is used in human clinical diagnostics. Forensic genetic laboratories have also investigated NGS technology, and there has been a small increase in the number of scientific articles and conference presentations dealing with forensic aspects of NGS in recent years. These contributions have been assessed as offering new possibilities for forensic genetic casework. More information can be obtained from processing the samples in a single experiment by applying combinations of markers (short tandem repeats (STRs), single nucleotide polymorphisms (SNPs), insertions/deletions, and messenger RNA (mRNA)) that cannot be analyzed simultaneously with today's standard polymerase chain reaction-capillary electrophoresis (PCR-CE) methods (Børsting & Morling, 2015).

DNA sequencing has a long history in forensic genetics. In the early 1990s, STRs were introduced as polymorphic DNA loci in the forensic field (Puers et al., 1993) and have become the primary workhorse for individual identification in criminal cases, paternity analyses, and missing person identification (Gill et al., 1997). Two seminal articles describing methods for DNA sequencing were published in 1977. Allan Maxam and Walter Gilbert described a method for separating base-specific chemical cleavage

Next Generation Sequencing (NGS) Technology in DNA Analysis
ISBN 978-0-323-99144-5, https://doi.org/10.1016/B978-0-323-99144-5.00023-8
© 2024 Elsevier Inc.
All rights reserved.

products from terminally labeled DNA fragments using gel electrophoresis (Maxam & Gilbert, 1977; Voelkerding et al., 2009). The Sanger dideoxynucleotide (ddNTP) chain terminating method was used for sequencing, in which the addition of a ddNTP to a growing DNA chain prevented further extension by the DNA polymerase (Sanger et al., 1977). This method was considered the gold standard for nucleic acid sequencing for the subsequent two and a half decades (Grada & Weinbrecht, 2013). Initially, the synthesized DNA fragments were separated by slab gel electrophoresis and detected by incorporating radioactively or fluorescently labeled deoxynucleotides (dNTPs) into the DNA fragments. Following that, fluorescently labeled ddNTPs and CE platforms were introduced, with increased sensitivity and throughput and reduced the cost of Sanger sequencing to the point where complete genome sequencing became feasible (Mitchelson, 2003). Sanger technology was used in an industrial, high-throughput configuration to sequence the first human genome, which was completed in 2003 as part of the Human Genome Project, a 13-year effort with an estimated cost of $2.7 billion (Wheeler et al., 2008).

Specific guidelines were adapted to select the proper STR loci used in forensic DNA analysis. Thus, core loci were defined, with significant overlap between international legislation (Welch et al., 2012). PCR-based amplicon sizing methods and CE systems were used to identify allele categories (Gill et al., 1997) using a simple nomenclature convention (Butler, 2006). However, the Sanger sequencing method was used continuously for verification and identification of, e.g., STR alleles (see references in STRbase, http://www.cstl.nist.gov/strbase/).

There is a continuous need to use CE systems. However, in terms of captured information, multiplex sizes, and analyzing highly degraded samples, NGS is adding a new dimension to the field of forensic genetics and offering distinct advantages over CE systems (Børsting & Morling, 2015). The biggest advantage is that STR-generated data in NGS is also compatible with any CE system-based database (Scheible et al., 2014). Additionally, NGS-derived STR genotypes supplement CE-derived genotypes by capturing the full nucleotide sequence underlying the repeat units and nearby flanking regions. Forensic tests using NGS will be able to distinguish STR variants that cannot be distinguished by mass spectrometry, such as repeat motifs that are shifted relative to each other in the repeat region (Pitterl et al., 2008). NGS STR typing demonstrates that it will be extremely useful in routine casework by increasing discrimination power, improving mixture resolution, and improving the identification of stutter peaks and artifacts (Gettings et al., 2015). The true variation in core forensic STR loci has been discovered, as have previously unknown STR alleles (Børsting & Morling, 2015). Although NGS has largely replaced conventional Sanger sequencing in genomic research, it has not yet made its way into normal practice (Behjati & Tarpey, 2013). Also, NGS STR analysis presents challenges for forensic practitioners. The new technology will have an impact on how data is analyzed and reported, as well as how it is stored and searched in databases (Parson et al., 2016).

DNA sequencing platforms

NGS platforms all have one thing in common: MPS of clonally amplified or single DNA molecules separated by space in a flow cell. This design is different from Sanger sequencing, which is based on the electrophoretic separation of chain-termination products generated in individual sequencing reactions. Sequencing in NGS is accomplished through repeated cycles of polymerase-mediated nucleotide extensions or, in one format, iterative cycles of oligonucleotide ligation. Depending on the platform, NGS generates hundreds of megabases to gigabases of nucleotide-sequence output in a single instrument run (Voelkerding et al., 2009).

In 1996, pyrosequencing was introduced as a real-time sequencing alternative to Sanger sequencing (Ronaghi et al., 1996). The nucleotides were added sequentially to the DNA synthesis reaction, and the pyrophosphate produced was used to generate light via a cascade of enzymatic reactions involving the three enzymes ATP Sulfurylase, Luciferase, and Apyrase. Because the light was detected in real-time by a CCD camera, no electrophoresis of the sequencing products was required. Pyrosequencing was less expensive and faster than Sanger sequencing, and the method was applied to mtDNA sequencing (Andréasson et al., 2002) and later also used for STR sequencing (Divne et al., 2010). However, the short sequencing length and especially the limited multiplexing capability of the instruments were not compatible with the low amounts of DNA usually recovered from trace samples, and the method was never used in casework (Divne et al., 2010). The first commercial high-throughput sequencing platform, the Genome Sequencer 20 from 454 Life Sciences, used pyrosequencing (Margulies et al., 2005). Despite the fact that the first pyrosequencing instruments were never widely used in science, with this technology, it was possible to sequence the human genome in 5 months for $1.5 million (Wheeler et al., 2008). Since then, several high-throughput sequencing methods and platforms have been introduced. The majority of them have been acquired by larger corporations, and the instruments' names have occasionally changed; for example, Solexa was renamed Illumina. Some have come and gone, such as Helicos BioSciences' HeliScope platform (Braslavsky et al., 2003), and Roche announced in 2015 that the highly successful 454 pyrosequencing would be phased out. The Supported Oligonucleotide Ligation and Detection System 2.0 platform, which is distributed by Applied Biosystems, was developed in the laboratory of George Church and reported in 2005 along with the resequencing of the *Escherichia coli* genome (Shendure et al., 2005). There are countless numbers of commercial capture assays available from different companies, e.g., the SureSelect Human All Exon Kit (Agilent), the HaloPlex Exome Kit (Agilent), the Ion AmpliSeq Exome RDY Kit (Life Technologies), the Nextera Rapid Capture Exome Kit (Illumina), the SeqCap EZ Human Exome Library (Roche NimbleGen), the TruSight One Sequencing Panel (Illumina) that targets 4800 genes, or the Ion AmpliSeq Cancer Panel (Life Technologies) that targets 400 genes

by PCR. The above-mentioned companies also provide services for the generation of customized panels defined by the user for specific projects or purposes (Børsting & Morling, 2015).

Overview of the methodology

Although each NGS technology has its own approach to sequencing, the NGS platforms share a common foundation of template preparation, sequencing and imaging, and data analysis (Metzker, 2010).

Template preparation

The process of template preparation entails creating a nucleic acid library (DNA or complementary DNA (cDNA)) and amplifying it. The DNA (or cDNA) sample is fragmented, and adapter sequences (synthetic oligonucleotides of a known sequence) are ligated onto the ends of the DNA fragments to create sequencing libraries. Libraries are clonally amplified in preparation for sequencing after they have been produced (Berglund et al., 2011; Quail et al., 2012).

Sequencing and imaging

In order to retrieve nucleic acid sequences from amplified libraries, most platforms rely on synthesis by sequencing. The library fragments serve as a template for the creation of a new DNA fragment. Washing and flooding the fragments with the known nucleotides in a sequential order is used to sequence them. Nucleotides are digitally recorded as sequences as they are incorporated into the growing DNA strand. For detecting nucleotide sequence information, the platforms rely on a slightly different mechanism for detecting nucleotide sequence information. For example, the Life Technologies Ion Torrent Personal Genome Machine (PGM) uses semiconductor sequencing, which detects pH changes caused by the release of a hydrogen ion as a nucleotide is incorporated into a developing strand of DNA (Quail et al., 2012). The MiSeq, on the other hand, depends on the detection of fluorescence caused by the integration of fluorescently labeled nucleotides into the expanding DNA strand (Quail et al., 2012).

Data analysis

Raw sequence data must go through numerous analysis procedures after it has been sequenced. Preprocessing the data to remove adapter sequences and low-quality reads, mapping the data to a reference genome or *de novo* alignment of the sequence reads, and analyzing the collected sequence are all part of a generic data analysis workflow

for NGS data. The sequence analysis can include a wide range of bioinformatics assessments, such as genetic variants requiring the detection of SNPs or indels (i.e., the insertion or deletion of bases), the discovery of novel genes or regulatory elements, and the assessment of transcript expression levels. The identification of both somatic and germline mutation events that may contribute to the diagnosis of a disease or genetic condition can also be included in the analysis. There are numerous free online tools and software packages available to perform the bioinformatics required to fully analyze sequence data (Gogol-Döring & Chen, 2012).

FDSTools, an open-source software program created especially for this purpose, can be used to examine, decipher, and summarize forensic MPS-STR data. It can offer a solution that would allow us to not only process all raw MPS data but also to have a straightforward, portable tool that would enable a DNA specialist to summarize, illustrate, and flexibly explain nearly all individual DNA sequences in court. When MPS is employed, STRs are assessed as sequence variations with distinct, precisely determinable stutter features (Hoogenboom et al., 2017).

For each individual allele, FDSTools analyzes a database of reference samples to identify stuttering, and other systemic PCR or sequencing errors. Moreover, stutter models are developed for every repeating element to forecast stutter artifacts for alleles that are not represented in the reference set. Following that, the noise in a sequence profile is identified, and any necessary adjustments are made. The end result is a more accurate depiction of a sample's actual composition (Hoogenboom et al., 2017). Understanding the experimental error profile of the entire analytical technique is essential for a thorough interpretation of each experiment's findings. In contrast to CE, where one "only" needs to monitor problems like peak-bleed through, peak shifts, allele imbalance, unusually high stutters, and new alleles at unexpected locations in the STR profile, MPS requires one to investigate the sequence variation among many millions of individual sequence reads. Similar to CE, where peaks are typically only accepted as genuine peaks if they exceed a predetermined fluorescent intensity detection threshold (let's say 50 RFUs), with MPS, at least when using FDSTools, one must also set an analytical threshold, or AT, in this case, the quantity of reads with the same sequence structure. The experimental design has a significant impact. One would anticipate fewer readings per sample and per locus (in the case of a multiplex STR design) in a run with many different DNA samples pooled for database purposes than one with fewer samples pooled.

The impact of NGS on genomics research

NGS has significantly accelerated multiple areas of genomics research in the previous years since the first commercial platform became available, enabling experiments that were previously not technically feasible or affordable (Voelkerding et al., 2009). The applications of NGS appear to be nearly limitless, allowing for rapid advances in many fields

related to the biological sciences. The human genome is being resequenced in order to identify genes and regulatory elements involved in pathological processes (Grada & Weinbrecht, 2013). Through whole-genome sequencing of a wide range of organisms, NGS has also provided a wealth of information for comparative biology studies. NGS is used in the fields of public health and epidemiology to identify novel virulence factors by sequencing bacterial and viral species. As NGS becomes more popular, it is inevitable that new and innovative applications will emerge.

Genomic analysis

NGS's high-throughput capability has been used to sequence entire genomes ranging from microbes to humans (Margulies et al., 2005; Pearson et al., 2007). NGS can be used to sequence entire genomes or be constrained to specific areas of interest, including all 22,000 coding genes (a whole exome) or small numbers of individual genes (Behjati & Tarpey, 2013). The capacity to sequence full human genomes at a low cost using NGS has sparked an international effort to sequence thousands of human genomes over the next decade (http://www.1000genomes.org), allowing for unprecedented characterization and cataloging of human genetic variation (Voelkerding et al., 2009).

Targeted genomic resequencing

Sequencing of genomic subregions and gene sets is being used to identify polymorphisms and mutations in cancer-related genes and regions of the human genome implicated in disease by linkage and whole-genome association studies (Voelkerding et al., 2009; Yeager et al., 2008). Particularly in the latter case, regions of interest can range from hundreds of kb (Almalki et al., 2017; Voelkerding et al., 2009; Yeager et al., 2008) to several Mb in size. Several genomic-enrichment steps, both traditional and novel, are being incorporated into overall experimental designs to maximize the use of NGS for sequencing such candidate regions (Voelkerding et al., 2009; Yeager et al., 2008). Although whole-genome and whole-exome sequencing are possible, targeted sequencing of specific genes or genomic regions is preferred in many cases where a suspected disease or condition has been identified. Targeted sequencing is less expensive, provides significantly more coverage of genomic regions of interest, and reduces sequencing cost and time (Xuan et al., 2013). Researchers have started to create sequencing panels that target hundreds of genomic regions known to be hotspots for disease-causing mutations. These panels sequence only desired regions of the genome, excluding the majority of the genome from analysis (Grada & Weinbrecht, 2013). Researchers and clinicians can create targeted sequencing panels that include specific genomic regions of interest. Furthermore, sequencing panels that target common regions of interest, such as hotspots for cancer-causing mutations, can be purchased for clinical

use (Rehm, 2013). Targeted sequencing, whether of single genes or entire panels of genomic regions, aids in the rapid diagnosis of many genetic diseases. The findings of disease-targeted sequencing can help with therapeutic decisions in a variety of diseases, including many cancers for which treatments can be cancer-type-specific (Rehm, 2013).

Metagenomics

The use of NGS has had a significant impact on the study of microbial diversity in environmental and clinical samples. In practice, genomic DNA is extracted from an interesting sample, converted to an NGS library, and sequenced. The output sequence is aligned to known reference sequences for microorganisms predicted to be present in the sample. Closely related species can be identified, and distantly related species can be deduced. Furthermore, *de novo* data set assembly can yield information to support the presence of known and potentially new species. Examples of metagenomic studies include the analysis of microbial populations in the ocean (Huber et al., 2007) and soil (Urich et al., 2008).

The impact of NGS on forensic genetics

The idea of sequencing every DNA (and/or RNA) molecule in the sample appeals to a forensic geneticist, who is used to dealing with the difficulty of obtaining sufficient information from trace samples, which frequently contain DNA from more than one contributor (Børsting & Morling, 2015). Today, the core forensic markers are typed with PCR-CE, and there are individual assays for autosomal STRs, Y-chromosome STRs, X-chromosome STRs, indels, mtDNA SNPs, autosomal SNPs, Y-chromosome SNPs, ancestry informative markers (AIMs), phenotypical markers, mRNA, etc. (de Knijff, 2019). One of the major advantages of NGS is that all (or most) of the PCR-CE assays can be combined into a single NGS assay if a capture for the relevant loci can be developed, such as in Precision ID GlobalFiler (Precision ID NGS STR Panel v2), which is compatible with the Ion S5/S5XL NGS platform from Applied Biosystems. Moreover, The ForenSeq DNA Signature Prep Kit (Verogen), formally released in 2015, has been extensively validated through a wide range of performance tests, including robustness, reproducibility, concordance with CE, and sensitivity of detection (Almalki et al., 2017; Churchill et al., 2016; Jäger et al., 2017; Köcher et al., 2018; Silvia et al., 2017). Additionally, there are specific kits that can process challenging samples in order to assess their applicability to genuine forensic cases, such as the DNA Signature Prep Kit, which has been used on numerous occasions (Almohammed et al., 2017; Köcher et al., 2018; Ma et al., 2016). In these tests, the kit outperformed CE on formalin-fixed paraffin-embedded tissue and bone samples dating from the 7th to 18th centuries, detecting a greater number of informative markers in seven out of 10 cases

(Churchill et al., 2016). The DNA Signature Prep Kit allows the most discriminating markers to be detected from limited samples, despite the level of DNA degradation (Khubrani et al., 2019; Votrubova et al., 2017). The increased discriminating power provided by combining STRs and SNPs will be useful in demanding cases such as mixture analysis, complex kinship cases, and degradation results of partial profiles (Khubrani et al., 2019; King et al., 2018).

NGS can be used to screen for variants in genes involved in the metabolism of specific drugs as well as to supplement toxicology investigations of a deceased person to determine whether an unexpected death was accidental or premeditated (VISSER et al., 2005). It will also be possible to examine DNA from bacteria, viruses, phages, and fungi from the deceased in order to identify disease-causing microorganisms or to look for microbial community imbalances that may provide clues to the cause of death (Cox et al., 2013). The sequencing of the microbiome in swabs or soil samples revealed significant differences in the taxa found in different locations (Giampaoli et al., 2014). This could be used to compare similarities between trace and reference samples. However, it should be noted that perfect matches, if not exclusions, are unlikely because the microbiome is constantly changing due to environmental factors such as temperature, humidity, sampling time, and so on. It was also discovered that samples were taken a few meters apart at the same time and only shared 50% of the microbiome diversity. Nevertheless, the variation between sampling sites was much higher (Young et al., 2014).

Short tandem repeats sequencing

Because of the large national DNA databases with STR profiles from criminal offenders and irreplaceable trace samples from old cases, STRs are essential to criminal casework and will continue to be so. As a result, any NGS assay intended for forensic genetics must be capable of sequencing the core STR loci. However, most NGS studies focus on SNPs, small indels, and copy number variations, while repeats have received little attention, despite the fact that repeats cover nearly half of the human genome and STRs alone account for 15% (Treangen & Salzberg, 2012). The discovery of numerous new STR and SNP-STR alleles with similar sizes renders the old PCR-CE-based STR allele nomenclature obsolete. In literature, Gelardi et al.'s (Gelardi et al., 2014) nomenclature was followed, which divides the name into four elements: (1) the locus name, (2) the length of the repeat region divided by the length of the repeat unit, (3) the sequence(s) of the repeat unit(s) followed by the number of repeats, and (4) variations in the flanking regions. Because NGS may identify more alleles, the implementation of NGS-based assays in forensic genetics necessitates a complete reevaluation of the current STR frequency databases. Another intriguing finding from the same study (Gelardi et al., 2014) was that approximately 30% of homozygous genotype calls made by PCR-CE turned out to be heterozygous when the individuals were sequenced. This demonstrates

yet another significant advantage of STR-NGS. If the contributors have alleles of the same size with different sequence compositions or if the true allele of the minor contributor has a different sequence than the stutter artifact of the major contributor, sequencing complex and compound STRs with many alleles of the same size may simplify mixture interpretation. It was recently demonstrated that NGS could detect minor contributor sequences in 1:100 and 1:50 mixtures (Fordyce et al., 2015), something that is impossible with the current PCR-CE technology. The reads from the minor contributor will be difficult to separate from stutters and noise sequences in these types of mixtures, but the fact that they can be identified opens up new possibilities in mixture interpretation, and it is certainly something that should be investigated further (Børsting & Morling, 2015).

The initial commercial NGS kits for forensic genetics

In 2014, Thermo Fisher Scientific released two SNP typing assays for the Ion PGM System: (1) the HID-Ion AmpliSeq Identity Panel for human identification (Børsting et al., 2014), which amplifies 124 autosomal SNPs, including the majority of the SNP*for*ID (Sanchez et al., 2006) and individual identification SNPs (IISNPs) (Pakstis et al., 2010); and 34 Y-chromosome SNPs, as well as (2) the HID-Ion AmpliSeq Ancestry Panel for ancestry estimation, which includes the majority of the AIMs from the Seldin (Nassir et al., 2009) and Kidd laboratory selection panels (Nievergelt et al., 2013). Thermo Fisher Scientific's strategy at the time was to develop assays that would be used in addition to PCR-CE typing. Illumina, on the other hand, had stated that their strategy would be to replace PCR-CE with PCR-NGS. The ForenSeq DNA Signature Prep Kit was developed in the fall of 2014 to amplify 27 autosomal STRs, 8 X-STRs, 25 Y-STRs, 95 autosomal human identification SNPs, 56 autosomal AIMs, and 24 autosomal SNPs associated with pigmentary traits in a single multiplex PCR. The multiplex includes, among others, all of the STR loci in the CODIS and European standard sets, most of the SNP*for*ID (Sanchez et al., 2006) and IISNPs (Pakstis et al., 2010) and all of the HIrisPlex loci (Walsh et al., 2013). The ForenSeq DNA Signature Prep Kit was introduced together with the MiSeq FGx platform, a MiSeq developed specifically for forensic genomics. One of the most difficult challenges will be creating a forensic NGS tool for analyzing and reporting sequence data. With NGS data, it is not possible to manually analyze the sequences or even each genotype call. As a result, before they can be used in criminal cases, software solutions must be completely trustworthy and thoroughly validated (Børsting & Morling, 2015). Software modules can be used to estimate bio-geographic ancestry, mtDNA haplogroups, Y-chromosome haplogroups, tissue identification, and phenotypes, among other things.

For the forensic community, commercially available MPS kits with PCR-based procedures targeting several loci in sizable multiplexes have been created, as discussed earlier.

These tests, however, might not be appropriate for forensic evidence with degraded DNA, like skeletonized remains and hair shafts. DNA damage from postmortem environmental exposures is known to occur in the form of base alteration such as depurination and cytosine deamination as well as fragmentation (Hofreiter et al., 2001). It has been demonstrated that DNA from chemically and age-treated bone samples is comparable to ancient DNA, with average lengths as low as 70 bp (Marshall et al., 2017).

Due to the DNA degradation, commercial forensic MPS tests, which typically require >100 bp fragments, may not be able to successfully enrich the sample. Thankfully, hybridization capture and MPS can be used to assess this and other damaged DNA samples with short fragments. Single-stranded DNA or RNA probes are used in the hybridization process to anneal or hybridize denatured DNA strands that are complementary to the probe sequence. These hybridized DNA strands are then collected while the unhybridized DNA is washed away. DNA fragments as short as 30−35 bp in length are susceptible to hybridization capture, making it possible to successfully profile DNA from even the most damaged forensic DNA samples (Gorden et al., 2021).

Current status of next generation sequencing

Few of the MPS-based results from forensic investigations have been presented in court. This shows that MPS is still not frequently applied in forensics (de Knijff, 2019). This is in direct opposition to other fields of genetic diagnostics like oncogenetics and clinical genetics, where tens of thousands of DNA samples have been regularly tested each year since 2010, utilizing either targeted genome sequencing or whole genome sequencing with MPS. There are, of course, a variety of explanatory factors for this. First, these other genetic fields have been using SNP array platforms to screen DNA samples for genetic variants at the entire genome level for at least 20 years without ever being constrained by the availability of usually minute amounts of degraded DNA (LaFramboise, 2009) using techniques for bioinformatics to analyze their findings. It was a relatively short step from there for these laboratories to transition into the more complicated MPS era. They could quickly profit from the already established and authoritative human genome variation society nomenclature criteria, and they were also quick to act in terms of ethical norms (den Dunnen et al., 2016). The disparity in the measured outcomes of the two technologies is the most evident and, in many ways, also the most significant distinction between MPS results and CE results. For the purpose of STR genotyping, CE converts machine-measured DNA molecule migration times into DNA fragment lengths. To make understanding easier, these fragment lengths are then represented in peak profiles and tables as a relatively straightforward string of numbers. The underlying base pair variation of the examined DNA sample is not disclosed by CE, though this has a major consequence: CE analysis of STRs (Kimpton et al., 1993) undervalues the genetic

variation that exists in the DNA sample at the genetic level. It is commonly recognized that homoplasy, which is the inability to detect similar-sized DNA fragments with varied sequence contents but identical fragment sizes in CE, exists. Using MPS, the final experimental result is expressed as DNA sequences that expose all underlying sequence variations in the targeted DNA sample, regardless of the underlying sequence technique. If one wants to compare MPS STR results with CE STR data, one can translate these DNA sequences, in the case of STRs, into DNA fragment lengths, but this is not necessarily necessary. Homoplasy is no longer a concern; genuine alleles and stutter alleles can be clearly recognized in DNA samples from a single source. Unfortunately, stutter alleles that cannot be separated from real minor contributor alleles frequently make it difficult to analyze unbalanced mixtures with low minor contributions. The incorrect base pair substitutions, primarily caused by DNA editing mistakes made during PCR, are likewise not identified by CE analysis of STRs since they have no effect on the underlying fragment lengths. In this way, MPS exposes the whole range of errors: (a) DNA slippage during PCR causes stutters; (b) DNA editing faults during PCR create base pair errors; (c) strand slippage (mostly at homopolymer stretches) during sequencing causes strand errors; and (d) substitution type miscalls during sequencing result in base pair errors (Schirmer et al., 2015). There is enough data to conclude that the latter two sources of error are platform-specific to the sequence. The output of STR CE analysis is then converted into a very basic data file that essentially simply provides the names of the STR loci, the length(s) of the STR alleles, and the fluorescence intensities (or peak heights) identified by the CE platform. Peak profiles, which have been in use for more than 20 years and are simple to explain but not necessarily easy to grasp, can be used to display these data (Kircher et al., 2009).

Two very straightforward file formats, FASTQ (Cock et al., 2009) and/or FASTA (Pearson & Lipman, 1988), each of which can hold all individual DNA sequence reads generated during the MPS analysis, can be used to store the results of an MPS experiment. Nevertheless, because MPS platforms generate between a few tens of millions and a few hundred million sequence reads in a single experiment, one must rely on specifically created software that converts these millions of reads into an experimental summary that one can comprehend and communicate. The MPS software packages used to interpret, summarize, and visualize all sequence data also need to be able to distinguish between them in such a way that, ideally, one can always, in retrospect, go back to the original sequence data and explore them in alternative ways if requested. This is because these DNA sequence files contain all reads produced by the platform, i.e., those representing true alleles and those containing any kind of PCR/sequence error. With MPS, there is no longer a gold standard in terms of the platform and software, in contrast to the CE analysis of STRs. As a result, it is more challenging to compare MPS results directly across platforms, laboratories, and forensic DNA experts. Also, and probably more importantly, new nomenclature standards that, ideally, have maximum clarity are

required in order to convert MPS STR results into a format that can be compared with CE STR results. In short, MPS data represent the whole range of potential DNA sequence variation, whereas CE results only tell us about DNA fragment length variation when used to investigate the STR variation in a crime scene sample. It may be tempting to convert intricate MPS STR results into something akin to CE STR results, but doing so would mean ignoring all further genetic variation data that would be essential for a criminal investigation (Hoogenboom et al., 2017). There may be legal restrictions in many countries that could prevent the reporting scientists from including all MPS results of a forensic investigation in the form of, for example, a FASTA file or the full description of all the sequence variations because, with MPS one obtains the full sequence structure of the DNA fragments that were targeted. These genetic records are undoubtedly available and have been added to the overall inquiry file, but both the prosecution and the defense must make separate requests to obtain these supporting records. When it comes to CE, this extra data is typically presented as peak profiles, electropherograms, or tables that list all the underlying STR data. As previously said, this is a little more complicated for MPS. There are also legal limitations on the precise locus types that can be used (and reported) in forensic DNA investigations (de Knijff, 2019). Only digitally recorded and only made available upon specific request, the underlying data for the MPS experiment is available in FASTQ and FASTA files and includes all error reads, an error read summary, and other pertinent data on the quality of the experiment. For presenting MPS-STR results in court, all data is available as HTML-format files that may be viewed and debated with the aid of a graphical user interface and any browser after being processed with FDSTools (Hoogenboom et al., 2017).

There is a lot more information that needs to be kept in addition to the genetic data, whether it be error readings or reads reflecting real alleles. It is common practice to combine many DNA samples for sequencing because a single run on an MPS machine generates millions of sequence reads. How many you can pool depends a lot on how many sequence reads are needed for each sample and each allele. A flexible case-by-case and/or sample-by-sample experimental strategy is now more viable as a result. It is quite easy to store the outcomes of STR CE genotyping in any STR profile database. The three default parameters that can be entered as extremely basic and condensed text strings are the sample code, locus name, and genotyping result. Individual peak heights and the multiplex STR kit used to create the profile are two examples of additional data that may be beneficial to include. Companies that sell multiplex STR genotyping kits take into account these guidelines because there is a universally recognized, consistent, standardized allele-calling nomenclature for CE. There are absolutely no standards for MPS. MPS can be executed on a variety of systems. While there have been a few recent initiatives, there are not yet any universal nomenclature rules (Phillips et al., 2018). Unless

one chooses to utilize an allele coding system like that used for the HLA system since 1968, it would still be exceedingly challenging to condense the entire range of sequence variation of every MPS discovered STR allele into a short and easy text string ("Nomenclature for Factors of the HL-a System," 1968). The primary benefit of an allele coding system like the one used with HLA is that the result of the STR sequence can be recorded as a relatively short allele designator.

Recommendations

Often, this entails a comprehensive set of suggestions or rules that offer standards for every conceivable technological, interpretative, and reporting difficulty. A few practical problems also need to be resolved, such as how to make the various National DNA databases accept STR alleles identified by MPS that include the complete range of genetic diversity found. Due to the large variety of MPS platforms and software that enable MPS outcomes, as opposed to CE, providing such recommendations will be more difficult because they must cover a much wider range of difficulties. At least the following concerns should be included in these, in no particular order (de Knijff, 2019):

1. A standardized nomenclature for MPS-based STR alleles that enables understanding/reconstructing the entire genetic variation spectrum without the necessity for back referencing, as in the case of the HLA nomenclature system.
2. Recommendations for the bare minimum of reads needed to accurately identify an STR allele in different scenarios, such as a straightforward reference database sample, a sample from a single crime scene, or a sample from a combination of multiple crime scenes.
3. There are three recommendations that can be utilized to give details on the whole range of nontarget (or incorrect) reads. This can also involve disclosing details regarding the sample pool method that was applied to the samples' screening as well as the documentation of barcode techniques.
4. Suggestions on the MPS approach that was employed: Was a less inclusive sequence technique employed, or were the entire reads 1 and 2 sequenced forward and reverse before being assembled and aligned? Was the PCR amplicon sequenced in its entirety, depending on the MPS platform, or were assembled, partially sequenced amplicons aligned?
5. Suggestions for the file formats required to hold all MPS results.
6. Software used to analyze and summarize MPS results must meet six minimum requirements. What details should be available right away?
7. It is necessary to adapt statistical software programs to the novel allele designations provided by MPS experiments in order to understand the evidentiary value of CE-based matches between STR profiles.

Concluding remarks

High-throughput sequencing has accelerated research in a wide range of biological and applied science fields. The use of NGS in forensic genetics has been debated in recent years, and we are now seeing applications aimed specifically at human identification and phenotypic trait determination. The cost of instruments and kits will determine how quickly the transition from CE to NGS takes place, as well as the proportion of cases investigated by NGS. Other critical aspects of incorporating NGS into forensic genetics include the development and validation of software solutions.

References

Almalki, N., Chow, H. Y., Sharma, V., Hart, K., Siegel, D., & Wurmbach, E. (2017). Systematic assessment of the performance of illumina's MiSeq FGx™ forensic genomics system. *Electrophoresis, 38*(6), 846–854. https://doi.org/10.1002/elps.201600511

Almohammed, E., Zgonjanin, D., Iyengar, A., Ballard, D., Devesse, L., & Sibte, H. (2017). A study of degraded skeletal samples using ForenSeq DNA Signature™ Kit. *Forensic Science International: Genetics Supplement Series, 6*, e410–e412. https://doi.org/10.1016/j.fsigss.2017.09.158

Andréasson, H., Asp, A., Alderborn, A., Gyllensten, U., & Allen, M. (2002). Mitochondrial sequence analysis for forensic identification using pyrosequencing technology. *Biotechniques, 32*(1), 124–133. https://doi.org/10.2144/02321rr01

Behjati, S., & Tarpey, P. S. (2013). What is next generation sequencing? *Archives of Disease in Childhood: Education and Practice Edition, 98*(6), 236–238. https://doi.org/10.1136/archdischild-2013-304340

Berglund, E. C., Kiialainen, A., & Syvänen, A. C. (2011). Next-generation sequencing technologies and applications for human genetic history and forensics. *Investigative Genetics, 2*(1). https://doi.org/10.1186/2041-2223-2-23

Børsting, C., Fordyce, S. L., Olofsson, J., Mogensen, H. S., & Morling, N. (2014). Evaluation of the ion Torrent™ HID SNP 169-plex: A SNP typing assay developed for human identification by second generation sequencing. *Forensic Science International: Genetics, 12*, 144–154. https://doi.org/10.1016/j.fsigen.2014.06.004

Børsting, C., & Morling, N. (2015). Next generation sequencing and its applications in forensic genetics. *Forensic Science International: Genetics, 18*, 78–89. https://doi.org/10.1016/j.fsigen.2015.02.002

Braslavsky, I., Hebert, B., Kartalov, E., & Quake, S. R. (2003). Sequence information can be obtained from single DNA molecules. *Proceedings of the National Academy of Sciences, 100*(7), 3960–3964. https://doi.org/10.1073/pnas.0230489100

Butler, J. M. (2006). Genetics and genomics of core short tandem repeat loci used in human identity testing. *Journal of Forensic Sciences, 51*(2), 253–265. https://doi.org/10.1111/j.1556-4029.2006.00046.x

Churchill, J. D., Schmedes, S. E., King, J. L., & Budowle, B. (2016). Evaluation of the Illumina® beta version ForenSeq™ DNA signature Prep kit for use in genetic profiling. *Forensic Science International: Genetics, 20*, 20–29. https://doi.org/10.1016/j.fsigen.2015.09.009

Cock, P. J. A., Fields, C. J., Goto, N., Heuer, M. L., & Rice, P. M. (2009). The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. *Nucleic Acids Research, 38*(6), 1767–1771. https://doi.org/10.1093/nar/gkp1137

Cox, M. J., Cookson, W. O. C. M., & Moffatt, M. F. (2013). Sequencing the human microbiome in health and disease. *Human Molecular Genetics, 22*(1), R88–R94. https://doi.org/10.1093/hmg/ddt398

Divne, A. M., Edlund, H., & Allen, M. (2010). Forensic analysis of autosomal STR markers using pyrosequencing. *Forensic Science International: Genetics, 4*(2), 122–129. https://doi.org/10.1016/j.fsigen.2009.07.004

den Dunnen, J. T., Dalgleish, R., Maglott, D. R., Hart, R. K., Greenblatt, M. S., Mcgowan-Jordan, J., Roux, A. F., Smith, T., Antonarakis, S. E., & Taschner, P. E. M. (2016). HGVS recommendations

for the description of sequence variants: 2016 update. *Human Mutation, 37*(6), 564–569. https://doi.org/10.1002/humu.22981

Fordyce, S. L., Mogensen, H. S., Børsting, C., Lagacé, R. E., Chang, C. W., Rajagopalan, N., & Morling, N. (2015). Second-generation sequencing of forensic STRs using the ion Torrent™ HID STR 10-plex and the ion PGM™. *Forensic Science International: Genetics, 14*, 132–140. https://doi.org/10.1016/j.fsigen.2014.09.020

Gelardi, C., Rockenbauer, E., Dalsgaard, S., Børsting, C., & Morling, N. (2014). Second generation sequencing of three STRs D3S1358, D12S391 and D21S11 in danes and a new nomenclature for sequenced STR alleles. *Forensic Science International: Genetics, 12*, 38–41. https://doi.org/10.1016/j.fsigen.2014.04.016

Gettings, K. B., Aponte, R. A., Vallone, P. M., & Butler, J. M. (2015). STR allele sequence variation: Current knowledge and future issues. *Forensic Science International: Genetics, 18*, 118–130. https://doi.org/10.1016/j.fsigen.2015.06.005

Giampaoli, S., Berti, A., Di Maggio, R. M., Pilli, E., Valentini, A., Valeriani, F., Gianfranceschi, G., Barni, F., Ripani, L., & Romano Spica, V. (2014). The environmental biological signature: NGS profiling for forensic comparison of soils. *Forensic Science International, 240*, 41–47. https://doi.org/10.1016/j.forsciint.2014.02.028

Gill, P., Sparkes, R., & Kimpton, C. (1997). Development of guidelines to designate alleles using an STR multiplex system. *Forensic Science International, 89*(3), 185–197. https://doi.org/10.1016/S0379-0738(97)00131-X

Gogol-Döring, A., & Chen, W. (2012). An overview of the analysis of next generation sequencing data. *Methods in Molecular Biology, 802*, 249–257. https://doi.org/10.1007/978-1-61779-400-1_16

Gorden, E. M., Sturk-Andreaggi, K., & Marshall, C. (2021). Capture enrichment and massively parallel sequencing for human identification. *Forensic Science International: Genetics, 53*. https://doi.org/10.1016/j.fsigen.2021.102496

Grada, A., & Weinbrecht, K. (2013). Next-generation sequencing: Methodology and application. *Journal of Investigative Dermatology, 133*(8), e11. https://doi.org/10.1038/jid.2013.248

Hofreiter, M., Jaenicke, V., Serre, D., Von Haeseler, A., & Pääbo, S. (2001). DNA sequences from multiple amplifications reveal artifacts induced by cytosine deamination in ancient DNA. *Nucleic Acids Research, 29*(23), 4793–4799. https://doi.org/10.1093/nar/29.23.4793

Hoogenboom, J., van der Gaag, K. J., de Leeuw, R. H., Sijen, T., de Knijff, P., & Laros, J. F. J. (2017). FDSTools: A software package for analysis of massively parallel sequencing data with the ability to recognise and correct STR stutter and other PCR or sequencing noise. *Forensic Science International: Genetics, 27*, 27–40. https://doi.org/10.1016/j.fsigen.2016.11.007

Huber, J. A., Mark Welch, D. B., Morrison, H. G., Huse, S. M., Neal, P. R., Butterfield, D. A., & Sogin, M. L. (2007). Microbial population structures in the deep marine biosphere. *Science, 318*(5847), 97–100. https://doi.org/10.1126/science.1146689

Jäger, A. C., Alvarez, M. L., Davis, C. P., Guzmán, E., Han, Y., Way, L., Walichiewicz, P., Silva, D., Pham, N., Caves, G., Bruand, J., Schlesinger, F., Pond, S. J. K., Varlaro, J., Stephens, K. M., & Holt, C. L. (2017). Developmental validation of the MiSeq FGx forensic genomics system for targeted next generation sequencing in forensic DNA casework and database laboratories. *Forensic Science International: Genetics, 28*, 52–70. https://doi.org/10.1016/j.fsigen.2017.01.011

Khubrani, Y. M., Hallast, P., Jobling, M. A., & Wetton, J. H. (2019). Massively parallel sequencing of autosomal STRs and identity-informative SNPs highlights consanguinity in Saudi Arabia. *Forensic Science International: Genetics, 43*. https://doi.org/10.1016/j.fsigen.2019.102164

Kimpton, C. P., Gill, P., Walton, A., Urquhart, A., Millican, E. S., & Adams, M. (1993). Automated DNA profiling employing multiplex amplification of short tandem repeat loci. *Genome Research, 3*(1), 13–22. https://doi.org/10.1101/gr.3.1.13

King, J. L., Churchill, J. D., Novroski, N. M. M., Zeng, X., Warshauer, D. H., Seah, L. H., & Budowle, B. (2018). Increasing the discrimination power of ancestry- and identity-informative SNP loci within the ForenSeq™ DNA Signature Prep Kit. *Forensic Science International: Genetics, 36*, 60–76. https://doi.org/10.1016/j.fsigen.2018.06.005

Kircher, M., Stenzel, U., & Kelso, J. (2009). Improved base calling for the illumina genome analyzer using machine learning strategies. *Genome Biology, 10*(8). https://doi.org/10.1186/gb-2009-10-8-r83

de Knijff, P. (2019). From next generation sequencing to now generation sequencing in forensics. *Forensic Science International: Genetics, 38*, 175–180. https://doi.org/10.1016/j.fsigen.2018.10.017

Köcher, S., Müller, P., Berger, B., Bodner, M., Parson, W., Roewer, L., & Willuweit, S. (2018). Inter-laboratory validation study of the ForenSeq™ DNA signature Prep kit. *Forensic Science International: Genetics, 36*, 77–85. https://doi.org/10.1016/j.fsigen.2018.05.007

LaFramboise, T. (2009). Single nucleotide polymorphism arrays: A decade of biological, computational and technological advances. *Nucleic Acids Research, 37*(13), 4181–4193. https://doi.org/10.1093/nar/gkp552

Ma, Y., Kuang, J. Z., Nie, T. G., Zhu, W., & Yang, Z. (2016). Next generation sequencing: Improved resolution for paternal/maternal duos analysis. *Forensic Science International: Genetics, 24*, 83–85. https://doi.org/10.1016/j.fsigen.2016.05.015

Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A., Berka, J., Braverman, M. S., Chen, Y. J., Chen, Z., Dewell, S. B., Du, L., Fierro, J. M., Gomes, X. V., Godwin, B. C., He, W., Helgesen, S., Ho, C. H., Irzyk, G. P., … Rothberg, J. M. (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature, 437*(7057), 376–380. https://doi.org/10.1038/nature03959

Marshall, C., Sturk-Andreaggi, K., Daniels-Higginbotham, J., Oliver, R. S., Barritt-Ross, S., & McMahon, T. P. (2017). Performance evaluation of a mitogenome capture and illumina sequencing protocol using non-probative, case-type skeletal samples: Implications for the use of a positive control in a next-generation sequencing procedure. *Forensic Science International: Genetics, 31*, 198–206. https://doi.org/10.1016/j.fsigen.2017.09.001

Maxam, A. M., & Gilbert, W. (1977). A new method for sequencing DNA. *Proceedings of the National Academy of Sciences, 74*(2), 560–564. https://doi.org/10.1073/pnas.74.2.560

Metzker, M. L. (2010). Sequencing technologies the next generation. *Nature Reviews Genetics, 11*(1), 31–46. https://doi.org/10.1038/nrg2626

Mitchelson, K. R. (2003). The use of capillary electrophoresis for DNA polymorphism analysis. *Molecular Biotechnology, 24*(1), 41–68. https://doi.org/10.1385/MB:24:1:41

Nassir, R., Kosoy, R., Tian, C., White, P. A., Butler, L. M., Silva, G., Kittles, R., Alarcon-Riquelme, M. E., Gregersen, P. K., Belmont, J. W., De La Vega, F. M., & Seldin, M. F. (2009). An ancestry informative marker set for determining continental origin: Validation and extension using human genome diversity panels. *BMC Genetics, 10*. https://doi.org/10.1186/1471-2156-10-39

Nievergelt, C. M., Maihofer, A. X., Shekhtman, T., Libiger, O., Wang, X., Kidd, K. K., & Kidd, J. R. (2013). Inference of human continental origin and admixture proportions using a highly discriminative ancestry informative 41-SNP panel. *Investigative Genetics, 4*(1). https://doi.org/10.1186/2041-2223-4-13

Nomenclature for factors of the HL-a system. (1968). *Bulletin of the World Health Organization, 39*(3), 483–486.

Pakstis, A. J., Speed, W. C., Fang, R., Hyland, F. C. L., Furtado, M. R., Kidd, J. R., & Kidd, K. K. (2010). SNPs for a universal individual identification panel. *Human Genetics, 127*(3), 315–324. https://doi.org/10.1007/s00439-009-0771-1

Parson, W., Ballard, D., Budowle, B., Butler, J. M., Gettings, K. B., Gill, P., Gusmão, L., Hares, D. R., Irwin, J. A., King, J. L., Knijff, P. D., Morling, N., Prinz, M., Schneider, P. M., Neste, C. V., Willuweit, S., & Phillips, C. (2016). Massively parallel sequencing of forensic STRs: Considerations of the DNA commission of the International Society for Forensic Genetics (ISFG) on minimal nomenclature requirements. *Forensic Science International: Genetics, 22*, 54–63. https://doi.org/10.1016/j.fsigen.2016.01.009

Pearson, B. M., Gaskin, D. J. H., Segers, R. P. A. M., Wells, J. M., Nuijten, P. J. M., & Van Vliet, A. H. M. (2007). The complete genome sequence of campylobacter jejuni strain 81116 (NCTC11828). *Journal of Bacteriology, 189*(22), 8402–8403. https://doi.org/10.1128/JB.01404-07

Pearson, W. R., & Lipman, D. J. (1988). Improved tools for biological sequence comparison. *Proceedings of the National Academy of Sciences, 85*(8), 2444–2448. https://doi.org/10.1073/pnas.85.8.2444

Phillips, C., Gettings, K. B., King, J. L., Ballard, D., Bodner, M., Borsuk, L., & Parson, W. (2018). "The devil's in the detail": Release of an expanded, enhanced and dynamically revised forensic STR Sequence Guide. *Forensic Science International: Genetics, 34*, 162–169. https://doi.org/10.1016/j.fsigen.2018.02.017

Pitterl, F., Niederstätter, H., Huber, G., Zimmermann, B., Oberacher, H., & Parson, W. (2008). The next generation of DNA profiling - STR typing by multiplexed PCR - ion-pair RP LC-ESI time-of-flight MS. *Electrophoresis, 29*(23), 4739–4750. https://doi.org/10.1002/elps.200800209

Puers, C., Hammond, H. A., Jin, L., Caskey, C. T., & Schumm, J. W. (1993). Identification of repeat sequence heterogeneity at the polymorphic short tandem repeat locus HUMTH01[AATG](n) and reassignment of alleles in population analysis by using a locus-specific allelic ladder. *The American Journal of Human Genetics, 53*(4), 953–958.

Quail, M. A., Smith, M., Coupland, P., Otto, T. D., Harris, S. R., Connor, T. R., Bertoni, A., Swerdlow, H. P., & Gu, Y. (2012). A tale of three next generation sequencing platforms: Comparison of ion torrent, pacific biosciences and illumina MiSeq sequencers. *BMC Genomics, 13*(1). https://doi.org/10.1186/1471-2164-13-341

Rehm, H. L. (2013). Disease-targeted sequencing: A cornerstone in the clinic. *Nature Reviews Genetics, 14*(4), 295–300. https://doi.org/10.1038/nrg3463

Ronaghi, M., Karamohamed, S., Pettersson, B., Uhlén, M., & Nyrén, P. (1996). Real-time DNA sequencing using detection of pyrophosphate release. *Analytical Biochemistry, 242*(1), 84–89. https://doi.org/10.1006/abio.1996.0432

Sanchez, J. J., Phillips, C., Børsting, C., Balogh, K., Bogus, M., Fondevila, M., Harrison, C. D., Musgrave-Brown, E., Salas, A., Syndercombe-Court, D., Schneider, P. M., Carracedo, A., & Morling, N. (2006). A multiplex assay with 52 single nucleotide polymorphisms for human identification. *Electrophoresis, 27*(9), 1713–1724. https://doi.org/10.1002/elps.200500671

Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences*, 5463–5467. https://doi.org/10.1073/pnas.74.12.5463

Scheible, M., Loreille, O., Just, R., & Irwin, J. (2014). Short tandem repeat typing on the 454 platform: Strategies and considerations for targeted sequencing of common forensic markers. *Forensic Science International: Genetics, 12*, 107–119. https://doi.org/10.1016/j.fsigen.2014.04.010

Schirmer, M., Ijaz, U. Z., D'Amore, R., Hall, N., Sloan, W. T., & Quince, C. (2015). Insight into biases and sequencing errors for amplicon sequencing with the Illumina MiSeq platform. *Nucleic Acids Research, 43*(6). https://doi.org/10.1093/nar/gku1341

Shendure, J., Porreca, G. J., Reppas, N. B., Lin, X., McCutcheon, J. P., Rosenbaum, A. M., Wang, M. D., Zhang, K., Mitra, R. D., & Church, G. M. (2005). Accurate multiplex polony sequencing of an evolved bacterial genome. *Science, 309*(5741), 1728–1732. https://doi.org/10.1126/science.1117389

Silvia, A. L., Shugarts, N., & Smith, J. (2017). A preliminary assessment of the ForenSeq™ FGx system: Next generation sequencing of an STR and SNP multiplex. *International Journal of Legal Medicine, 131*(1), 73–86. https://doi.org/10.1007/s00414-016-1457-6

Treangen, T. J., & Salzberg, S. L. (2012). Repetitive DNA and next-generation sequencing: Computational challenges and solutions. *Nature Reviews Genetics, 13*(1), 36–46. https://doi.org/10.1038/nrg3117

Urich, T., Lanzén, A., Qi, J., Huson, D. H., Schleper, C., & Schuster, S. C. (2008). Simultaneous assessment of soil microbial community structure and function through analysis of the meta-transcriptome. *PLoS One, 3*(6). https://doi.org/10.1371/journal.pone.0002527

Visser, L., Vanschaik, R., Vanvliet, M., Trienekens, P., Desmet, P., Vulto, A., Hofman, A., Vanduijn, C., & Stricker, B. (2005). Allelic variants of cytochrome P450 2C9 modify the interaction between nonsteroidal anti-inflammatory drugs and coumarin anticoagulants. *Clinical Pharmacology & Therapeutics, 77*(6), 479–485. https://doi.org/10.1016/j.clpt.2005.02.009

Voelkerding, K. V., Dames, S. A., & Durtschi, J. D. (2009). Next-generation sequencing: from basic research to diagnostics. *Clinical Chemistry, 55*(4), 641–658. https://doi.org/10.1373/clinchem.2008.112789

Votrubova, J., Ambers, A., Budowle, B., & Vanek, D. (2017). Comparison of standard capillary electrophoresis based genotyping method and ForenSeq DNA Signature Prep kit (Illumina) on a set of challenging samples. *Forensic Science International: Genetics Supplement Series, 6*. https://doi.org/10.1016/j.fsigss.2017.09.027. e140–e142.

Walsh, S., Liu, F., Wollstein, A., Kovatsi, L., Ralf, A., Kosiniak-Kamysz, A., Branicki, W., & Kayser, M. (2013). The HIrisPlex system for simultaneous prediction of hair and eye colour from DNA. *Forensic Science International: Genetics,* 7(1), 98–115. https://doi.org/10.1016/j.fsigen.2012.07.005

Welch, L. A., Gill, P., Phillips, C., Ansell, R., Morling, N., Parson, W., Palo, J. U., & Bastisch, I. (2012). European network of forensic science institutes (ENFSI): Evaluation of new commercial STR multiplexes that include the European standard set (ESS) of markers. *Forensic Science International: Genetics, 6*(Issue 6), 819–826. https://doi.org/10.1016/j.fsigen.2012.03.005

Wheeler, D. A., Srinivasan, M., Egholm, M., Shen, Y., Chen, L., McGuire, A., He, W., Chen, Y. J., Makhijani, V., Roth, G. T., Gomes, X., Tartaro, K., Niazi, F., Turcotte, C. L., Irzyk, G. P., Lupski, J. R., Chinault, C., Song, X. Z., Liu, Y., … Rothberg, J. M. (2008). The complete genome of an individual by massively parallel DNA sequencing. *Nature, 452*(7189), 872–876. https://doi.org/10.1038/nature06884

Xuan, J., Yu, Y., Qing, T., Guo, L., & Shi, L. (2013). Next-generation sequencing in the clinic: Promises and challenges. *Cancer Letters, 340*(2), 284–295. https://doi.org/10.1016/j.canlet.2012.11.025

Yeager, M., Xiao, N., Hayes, R. B., Bouffard, P. P., Desany, B., Burdett, L., Orr, N., Matthews, C., Qi, L., Crenshaw, A., Markovic, Z., Fredrikson, K. M., Jacobs, K. B., Amundadottir, L., Jarvie, T. P., Hunter, D. J., Hoover, R., Thomas, G., Harkins, T. T., & Chanock, S. J. (2008). Comprehensive resequence analysis of a 136 kb region of human chromosome 8q24 associated with prostate and colon cancers. *Human Genetics, 124*(2), 161–170. https://doi.org/10.1007/s00439-008-0535-3

Young, J. M., Weyrich, L. S., & Cooper, A. (2014). Forensic soil DNA analysis using high-throughput sequencing: A comparison of four molecular markers. *Forensic Science International: Genetics, 13*, 176–184. https://doi.org/10.1016/j.fsigen.2014.07.014

CHAPTER 2

Using conventional STR technology in analyzing biological transfer evidence

J. Thomas McClintock
Liberty University, Department of Biology and Chemistry, Lynchburg, VA, United States

Introduction

Homecoming seemed to come early this year. The leaves had given way to the vibrant red and yellow fall colors; pep rallies were being held on campus; tailgating parties were popping up everywhere near the stadium; and an overall excitement seemed to fill the air on campus. After the game, there would be, no doubt, endless fraternity and sorority parties to celebrate the hometown football team's victory. Alpha Epsilon Pi, a major fraternity on campus, was hosting a party after the game, and everyone was expected to attend. Slightly past midnight, a female at the party, who was known around campus for her sexual promiscuity, started to solicit male companions presumably for sexual favors. After a short stint with one of the fraternity brothers, the female returned to the party and continued to consume alcohol. Several drinks later, she met another male and proceeded to return to the bedroom where she had been previously with the first male companion. However, by this time the female was quite inebriated, and after a short bit of "tossing" in the bed with the second male companion she passed out. Upon awakening in the early morning, the female noticed that her underwear was moist, that her clothing was disheveled, and that almost everyone, including the two males, had left the party. Not being able to recall what had happened with the two males the night before, and after some deliberations, the female was presented to the local hospital by midmorning and claimed that she had been raped. The sexual assault nurse examiner (SANE) questioned the young lady and collected evidentiary samples such as clothing (i.e., blouse, skirt, and underwear) and swabs from various areas of her body. The evidentiary samples, as well as a known reference sample, were sent to the state forensic laboratory for DNA analysis. Known reference samples were later collected from both of the male companions and also sent to the laboratory for DNA analysis. Following the completion of the DNA testing, the state forensic laboratory reported its findings to the state's attorney assigned to the case. In essence, the state laboratory reported that both males were minor contributors to the DNA profiles developed from the vaginal swabs of the victim, but a third major DNA profile was isolated from the underwear that was from an unidentified male. Upon further investigation and after additional testing, this DNA profile was determined to be from the fraternity brother whose bed

Next Generation Sequencing (NGS) Technology in DNA Analysis
ISBN 978-0-323-99144-5, https://doi.org/10.1016/B978-0-323-99144-5.00016-0
© 2024 Elsevier Inc.
All rights reserved.

and room were used by the female and two male companions on the night of the party. This third male was ultimately arrested and charged with first-degree rape.

The above case scenario is not an atypical situation or sexual assault charge. In this instance, the victim of such a sexual assault doesn't recall most of the events that occurred the night before except that there was more than one male involved who probably had consensual sex with her or by force. However, in this case, the male that was ultimately arrested and charged with rape was not involved in the sexual assault, but his DNA incriminated him simply because it had transferred from his bed to the victim's underwear. In this situation, all victim(s) should report such assaults to local authorities in an effort to identify the perpetrator(s) and find justice. This chapter will explore how such events/scenarios can occur and how forensic analysts need to be aware of such situations when the deposition of DNA is in question, especially in light of the sensitivity of today's technology and instrumentation and the ability to detect DNA from multiple sources.

Touch DNA (or trace DNA)

In "*an introduction to forensic DNA analysis*," Rudin and Inman (2002) define contamination "as the inadvertent addition of an individual's physiological material or DNA during or after collection of the sample as evidence …a contaminated sample is one in which the material was deposited during collection, preservation, handling, or analysis." When the addition of biological material is deliberately made to a sample, this constitutes evidence tampering. The unintentional addition of biological material during sample collection and handling constitutes contamination that could have a significant impact on the interpretation of the results and courtroom proceedings.

Forensic evidence needs to be free of foreign material, such as external DNA or other trace biological material for the testing process to present a true and accurate representation of the evidence and crime scene. Consequently, crime scene investigators and analysts that handle evidence are required to adhere to stringent protocols that require wearing protective clothing (i.e., gloves, masks, *etc.*) to avoid contaminating the evidence with shed hair and skin cells. Such protective "gear" protects the investigator/analyst from the sample (samples should always be handled as potentially infectious agents) and the sample from the handler (shed hair and sloughing of skin cells will contaminate the sample). Since DNA test methods are sensitive enough to generate a full DNA profile from as few as one to five shed skin or epithelial cells, great care in the handling of evidentiary samples is critical and must be maintained at all times.

The human body is constantly shedding dead skin cells—between 30,000 and 40,000 skin cells are shed every hour (Milestone, 2004). Thus, over a 24-h period, an individual may lose almost a million skin cells (Sender et al., 2016). Such shed epithelial cells may have nuclei (containing 23 pairs of chromosomes available for DNA testing) that are

visible in the cytoplasm of the cell or may have lost the cellular contents due to leakage or apoptosis (programmed cell death). In addition, several studies have demonstrated the presence of cell-free DNA from senescent cells, which can be used for DNA profiling (Anker et al., 2001; Botezatu et al., 2000; Garcia-Olmo et al., 2004). The nucleated cells, however, will contain DNA that can be useful for typing and linking an individual to a crime scene. In addition to the "normal" shedding process, the transfer of biological material from one object to another is influenced by both environmental (e.g., friction and pressure, and length of handling time) and genetic (e.g., "shedder" status) factors. These factors play a significant role in establishing the type and number of cells deposited on an individual (i.e., skin surface or clothing) or an inanimate object.

Since the introduction of DNA profiling, the analysis of DNA extracted from biological material has become a daily occurrence in forensic laboratories worldwide. Over the years, the technology to extract and analyze DNA has become very sophisticated. One such technology, called "touch DNA" or "contact trace DNA," refers to the DNA that is recovered from skin cells that are left behind when a person touches or comes into contact with items such as clothes, a weapon, or other objects. These epithelial cells can be collected using lifting tape, swabbed with a Q-tip, or even scraped from the clothes of the victim or objects and viewed microscopically and/or subjected to DNA analysis.

Direct and indirect transfer of biological evidence

Primary transfer. In the early 20th century, Edward Locard, a forensic scientist in France, theorized that the perpetrator of a crime would bring "something" to the crime scene and leave with "something" from the crime scene. This concept, known as Locard's exchange principle, leads to the statement that "every contact leaves a trace" and that such items that are left behind or picked up can be used as forensic evidence. Locard's exchange principle essentially states that when two objects come into contact with one another a transfer of material can occur. For example, when an individual picks up a pen epithelial cells or skin cells can be left on the touched item. This sort of transfer can also occur when an individual touches or comes into contact with another person. The transferred item, such as skin cells or hair, could then be analyzed for DNA. This type of exchange between an animate or inanimate object is called primary or direct transfer. Many other types of primary transfers have been documented. These include the transfer of saliva (which contains skin cells) when kissing a person, the exchange of skin cells during a handshake, or the transfer of skin cells on the waistband of a victim's underwear during a sexual assault. Keep in mind that all of these examples of primary or direct transfer involve the exchange of biological material that contains DNA.

Secondary transfer. Biological evidence can also be transferred through indirect contact. Indirect contact is when biological material or evidence is transferred to an animate or inanimate object and then transferred to another person or object. For example, if an

individual handled a pen or pencil and then another person picked up that pen or pencil, the second individual could pick up material or evidence from the first person who touched the object. Another example of indirect transfer is when hair or skin cells are left on a bed by one person but are then transferred to another (or second) person that comes into contact with the biological material. These examples of indirect transfers are also known as secondary transfers. In these examples, the DNA from the first person is transferred or exchanged and now can be detected on the second person (or object) that contacted the material or evidence. Imagine the legal consequences of such interactions. In essence, an individual could be identified from skin cells or DNA left on an important piece of crime scene evidence, accused of the crime, while never being present at the scene.

Tertiary transfer. Another type of transfer is called a tertiary transfer. For example, the perpetrator of rape leaves semen on the victim's bedsheet (direct or primary transfer). Then, the victim touches or comes into contact with semen on the bedsheet (indirect or secondary transfer). Finally, the victim wipes their hands or their body with a washcloth, and some of the semen transfers from their hands/body to the washcloth. Again, because of secondary and now tertiary transfer, biological evidence can be found in a location where the biological source of the evidence (i.e., the individual) was never present.

As demonstrated, biological material such as skin cells, semen, blood, saliva, and urine can be transferred from one object to another, analyzed, and lead to the arrest and conviction of the perpetrator. However, such a transfer of biological evidence could lead to wrongful convictions. As an example, in June 2012, Lukis Anderson, a homeless man, was charged with capital murder for the death of Raveesh Kumra, a wealthy multimillionaire from Silicon Valley. This charge was based on Anderson's DNA found at the crime scene. However, at the time of the murder, Anderson was hospitalized due to intoxication and in a nearly comatose condition, requiring medical attention and supervision. The question was, How did Anderson's DNA end up at the crime scene? During the investigation, Anderson's legal team learned that the paramedics that treated Anderson had actually responded to the murder scene of Kumra and had inadvertently transferred Anderson's DNA at the scene (Smith, 2016). This case of secondary transference is an example of the errors or risks that may occur when relying on DNA evidence. Moreover, such "misplaced" DNA evidence would, no doubt, lead to the wrongful conviction of an innocent person. An illustration of the primary, secondary, and tertiary transfer of biological material is shown in Fig. 2.1.

Effects of surface/substrate "material" and biological transfer. As demonstrated above, the transfer of biological evidence from an individual to an item or another person at a crime scene can happen in several ways. The amount of biological evidence or DNA transferred to a substrate depends on many factors, whereby the nature of the substrate (or object) is critical. Such findings would allow analysts to determine the "order" to analyze evidence

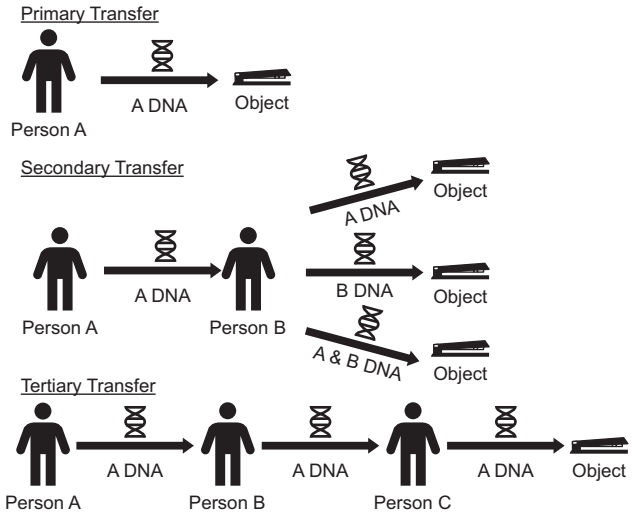


Figure 2.1. Illustration of primary, secondary, and tertiary transfer of biological material. *Source*: Evans, J., & Hadi, S. (2018). Low template DNA: Tad, touch, and traces. *Journal of Forensic Pathology, 3*, 1.

to save time and other valuable resources, especially in cases where a large number of evidentiary samples have been submitted for analysis. Daly et al. (2012) performed a study where three different substrates were selected to determine the amount of DNA that was transferred. The items or substrate materials in this study consisted of glass, fabric, and wood. These items were selected to represent objects such as bottles, firearms, clothing from a crime scene or sexual assault victim, and handles on knives. Volunteers handled the various items for 60 s; the DNA was recovered, quantified, extracted, amplified, and the profiles were analyzed and compared. Success rates for obtaining adequate DNA profiles of the handlers were approximately 36% for wood, 23% for fabric, and 9% for glass. These results suggest that the order of sampling or testing by the analysts in the laboratory should be surfaces made of wood, then clothing, and finally glass since substrates made of glass will have a low expectation of obtaining a meaningful DNA profile.

In another study, Fonnelop et al. (2015) demonstrated the transfer of DNA using various substrates. One part of this study investigated the primary transfer of touch DNA from plastic, wood, and metal surfaces. The results from this study demonstrated that DNA was readily transferred to wood and plastic, but less DNA was transferred to a metal substrate such as a door handle. These data support the earlier findings of Daly et al. that more DNA is transferred to wood than to glass, or in this instance, metal. Rough surfaces on the wood appeared to facilitate cell transfer and, thus, DNA, by the abrasive action. In addition, Fonnelop et al. were able to generate a full DNA profile after three transfers. Goray et al. (2010) also showed that the transfer of DNA was significantly greater when deposited on a porous substrate such as cotton than when deposited on a smooth, hard surface such as plastic. Lehmann et al. (2013) conducted a study where

a series of transfers of touch DNA were deposited on cotton or glass and then transferred onto "new" substrates, with six transfers in each series. A full DNA profile was observed on the first substrate, whereas a partial profile was only observed on the second to the fifth glass substrate. This study emphasized that great care and caution are needed when handling and examining evidentiary samples to prevent the transfer of biological material.

Effects on transfer and shedder status (assessment of shedder status and implications for secondary DNA transfer)

History of detecting biological transfer. In 1996, Kafarowski et al. initiated a study to evaluate the likelihood of the transfer of spermatozoa during laundering. At this time, there were minimum, if any, reports on the likelihood of spermatozoal transfer among laundered clothing items. Three couples participated in this study, whereby a semen stain was deposited on a clean pair of cotton panties by natural drainage following vaginal intercourse. The stains on the panties were air dried, marked or outlined for identification later, individually washed with three "pristine" pairs of cotton or cotton blend panties for a 10-min warm wash, followed by a cold rinse cycle with a phosphate-free detergent, and then machine dried. Following washing, the stains were examined for acid phosphatase (AP) activity, subjected to Christmas tree staining, and examined microscopically for the presence of spermatozoa. In all three instances, a trace amount of spermatozoa had transferred to the "pristine" panties during laundering. Of the 162 samples excised, 50% contained at least one sperm head, 38% contained one to two sperm heads, and 16% contained three to eight sperm heads. These observations revealed that spermatozoa could be present on a garment without any sexual activity involved and that these results could have strong repercussions with regard to an individual's testimony. Moreover, these results could be of significant importance when one considers cases involving victims/complainants who are not sexually active or who are unable to recall the details/specifics of the assault. In the absence of DNA testing, as observed in this study, transfer during washing warrants equal consideration with the transfer of biological material as a possible explanation for the presence of spermatozoa.

In 1999, Ladd et al. systematically evaluated the nature of biological transfer in an attempt to confirm the findings by Van Oorschot and Jones (1997) regarding the primary transfer of DNA from person to person or to various objects and the secondary transfer of DNA through an intermediary. In the first part of this study, individuals shook hands for various lengths of time (1, 5, 10, 30, and 60 s) and then handled precleaned items (e.g., door handles, telephone mouthpieces, computer keyboards, computer mice, and steering wheels) for 5 s. In the second part of the study, precleaned coffee mugs were handled for 2 h and then handled by a second individual for 10 s. Hands and objects were swabbed, the DNA extracted by organic extraction/microcon-purification, a method not commonly used today, quantified, and amplified using the AmpliType

PM and DQA1 typing kits (a kit no longer manufactured by Perkin Elmer Applied Biosystems) according to the manufacturer's protocol. In addition, biological transfer in this study was evaluated using short tandem repeats (STR) analysis using the AmpFlSTR Profiler Plus and COfiler DNA typing kits. The results showed that the primary transfer of DNA using the AmpliType PM and DQA1 typing systems could yield interpretable results, but detection was not guaranteed and appeared to be dependent on the individual performing the handshake or handling the objects. In many instances, there was insufficient DNA to generate an interpretable genotype, and since the smaller loci produced an amplified product versus the larger loci, it was determined that degradation appeared to be a factor.

The STR results were similar to those for the dot-blot-based (AmpliType PM and DQA1) systems. Degradation and low DNA yield appeared to be significant factors in locus and allelic dropout, with some allelic peaks barely exceeding background levels. In fact, a complete secondary STR profile was never detected. The authors concluded that "the primary transfer of DNA by handling is possible, but detecting an interpretable genotype is not assured." Moreover, secondary transfer was not observed; therefore, the data "do not support the inference that the interpretation of DNA profiles from case samples could be compromised by secondary transfer." When one considers today's level of improved extraction methods and purification yields, together with the increased level of instrumentation sensitivity, there is no wonder why this phenomenon has stimulated further investigation.

For all intense purposes, when a DNA profile is collected from an object at a crime scene, it is not possible to determine when the DNA was transferred to the object. To make matters worse, one cannot determine if this DNA profile was transferred prior to or during the course of the crime. After considering research performed by Van Oorschot et al. (1997) and Ladd et al. (1999), Lowe et al. (2002) demonstrated the difference between individuals in their tendency to deposit DNA during primary and secondary transfer, under ideal and specific laboratory conditions, to another individual and subsequently to an inanimate object. Specifically, the occurrence of a secondary transfer of DNA from one individual to another and then to an object can occur under favorable conditions. In fact, a full DNA profile from one individual was recovered from an item that they had not touched, "while the profile of the person having contact with the item was not observed." Interestingly, when a 30 min or 1 h delay was incorporated between human contact and contact with the object, both of the individuals' DNA profiles were detected. Interpretations of such transfers between individuals and inanimate objects agree with the findings of van Oorschot et al. (1999) as well as Ladd et al. (1999). However, interpretations of such results or DNA patterns may depend on the circumstances of the case and/or crime scene under investigation as well as the events that are described by the victim and/or suspect.

In 2012, Daly et al. examined the amount of DNA transferred from an individual to various inert surfaces such as glass, fabric, and wood. This study was discussed above, but the results demonstrated that there were differences in the amount of DNA transferred to these surfaces and that the success rate of achieving a full DNA profile could depend on the individual touching the object (i.e., a good vs. a bad shedder), the surface and nature of the object, and the activities leading up to the individual touching the object. Unfortunately, in casework, the forensic analyst will not have, in most instances, information about the individual's shedder status or the activity of the individual prior to touching or transferring biological material to an object.

As interest increased in understanding DNA transfer and its roles in criminal investigations, Goray et al. (2012) evaluated multiple transfers of DNA using mock crime scene scenarios. Although this study revealed mixed results, the authors concluded that the parameters of the experiment would have to be defined in order to align with the reported sequence of events. Such parameters would include, but not be limited to, type and duration of contact, type of substrate that the items (or samples) were transferred to, and environmental conditions such as temperature and humidity, together with other conditions that should be incorporated into the experimental design. The authors concluded that "such reenactment only becomes relevant and possible when clear and specific secondary or further transfer events are proposed during criminal investigation or court proceedings." The authors also suggest that a framework or database could be developed that comprises variables or factors previously studied that could lead to a predictive value as to their role in the transfer of biological material. As such, with the increasing role of DNA transfer used to explain the presence of specific biological material, it can be reasonable to expect that DNA experts will present and/or argue this line of defense as an explanation in courtroom proceedings.

STR analysis. Polymerase chain reaction (PCR)-based STR systems offer many advantages over earlier DNA typing techniques (e.g., restriction fragment length polymorphisms, or RFLP). STR systems provide a rapid and sensitive method to evaluate small amounts (1 ng) of human DNA. This small amount of DNA needed for STR systems is 50 times less than what is normally required for RFLP analysis. Also, the repeating sequences in an STR are relatively short, with the entire STR strand or allele generally beingless than 400 bp in length. This short length renders STR systems amendable to the analysis of samples suspected of being degraded. STR analysis often allows the DNA analyst to recover a complete DNA profile even from an evidentiary sample that was exposed to unfavorable conditions (e.g., body or stains subject to extreme decomposition). This is in sharp contrast to RFLP systems, which required large sample sizes for analysis and full-length fragments, which often consisted of thousands of bases, to generate a complete DNA profile.

More recent studies investigating the biological transfer of evidence have utilized STR analysis due to its level of sensitivity, its consistency, and the amount of variation

demonstrated between individuals. Human genomes contain 5%–10% of repetitive sequences that occur in tandem or adjacent to each other. These repetitive sequences, which appear more than once on the same chromosome, vary in size and length and show sufficient variability among individuals in a population. Regions of DNA that contain these short, repeated segments, or STRs, are important markers for human identity testing in the forensic community. There are literally thousands of STR markers scattered throughout the human genome, which occur, on average, one in every 10,000 nucleotides. The DNA sequence repeated in an STR motif is usually from 2 to 7 base pairs (bp), with 4 bases being the preferred size for forensic systems (Edwards et al. (1991, 1992; Warne et al. 1991). An example of a 4 bp or a tetranucleotide repeat is shown below (Fig. 2.2), where the TCTA motif is repeated 4 times.

STRs and their corresponding loci are easily amplified by PCR. Further, PCR amplification of many different STR loci is commonly performed simultaneously in the same tube. The simultaneous amplification of two or more loci is commonly known as multiplexing or multiplex PCR. Initially, for a multiplexing reaction to be successful, the system was designed to ensure that the size of the amplified products did not overlap, thereby allowing each STR allele for a specific locus to be clearly visualized on a gel or by capillary electrophoresis. However, this "requirement design" of overlapping fragments became less important with the development of multiple color detection systems.

Different detection methods are available to visualize the STR products. The STR loci and corresponding alleles may be separated by gel electrophoresis and detected using ethidium bromide, silver staining, or exotic dyes (e.g., SYBR green). Several STR systems have been developed where fluorescent dyes or labels are used to detect the STR alleles either during (i.e., capillary electrophoresis) or after separation (i.e., gel electrophoresis). The resulting STR profiles are routinely interpreted by direct comparison to DNA standards, allelic ladders (an artificial mixture of common genes present in the human population for a particular STR marker or locus), and reference standards (known DNA profiles from the victim and suspect). Probability calculations are determined based on classical population genetic principles.

For STR markers to be effective across various jurisdictions, a common set of standardized markers is used. Currently, the forensic scientific community in the United States has established a set of 13 core STR loci, which, in turn, can be entered into a national database known as the Combined DNA Index System (CODIS), a collection of DNA profiles from known offenders. A summary of the 13 CODIS loci is contained in Table 2.1.

.....ATGTGATCTATCTATCTATCTATTGG.....

Figure 2.2. A 4 bp (i.e., TCTA) or tetranucleotide repeat where the TCTA motif is repeated 4 times.

Table 2.1 Information on the 13 core short tandem loci listed in CODIS.

STR locus	Chromosome number	Sequence
FGA	4	CTTT
vWA	12	[TCTG][TCTA]
D3S1358	3	[TCTG][TCTA]
D21S11	21	[TCTA][TCTG]
D8S1179	8	TATC
D7S820	7	GATA
D13S317	13	TATC
D5S818	5	AGAT
D16S539	16	GATA
CSF1PO	5	AGAT
TPOX	2	AATG
THO1	11	TCAT
D18S51	18	AGAA

Persistence and transfer of DNA from laundered semen/biological stains. Sexual assaults account for more than 60% of criminally related cases. In fact, as of 1998, one out of every six American women has been the victim of an attempted or completed rape in her lifetime, and because the United States population has increased substantially since then, it is probable that the number of victims has as well (National Institute of Justice and Centers for Disease Control and Prevention). In most instances, these cases include a comprehensive case history, a SANE report, and biological samples such as swabs and clothing (e.g., underwear, blouses, pants, *etc.*) that are collected as part of the physical evidence recovery kit (PERK). Unfortunately, a number of sexual assault cases involve children and the elderly who are unable to provide an accurate and/or detailed account of the assault.

The identification of biological fluids, especially semen or spermatozoa, begins with a physical examination of the clothing and garments submitted in the PERK using an alternate light source. If presumed biological stains are observed it is imperative to determine the biological source of the evidence. If the stain is presumed to be semen, the analyst will test for the presence of acid phosphatase (AP) activity and/or the prostate-specific antigen (P30) using presumptive assays. If a significant amount of sample is present, a confirmatory test, such as the Christmas tree stain or the rapid stain identification for sperm, will be performed.

When semen is identified on the garments or underwear of the victim sexual activity is usually presumed. However, the manner of deposition may not be as evident as it appears. Situations when trace amounts of sperm are detected or the defendant is adamant about not having sexual contact with the complainant/victim can complicate the interpretations by the forensic analyst. In these instances, the question becomes the deposition of the semen on the clothing. In other words, was the semen deposited as a result of sexual activity or by an innocuous incident or source, such as the secondary transfer of

biological material during laundering? Moreover, the significance of these findings—in particular, any genetic profiles that are generated—may be explained in intrafamilial cases where biological material or DNA may be innocently deposited onto a victim's clothing by various means in a shared family household. This becomes even more relevant when considering today's sophisticated technology and instrumentation. For example, in 1996, Kafarowski et al. demonstrated that semen from a single pair of semen-stained panties could be transferred to previously unstained panties during laundering. In a later study, Noel et al. (2016) demonstrated that sufficient amounts of DNA could be transferred onto laundered clothing from bed sheets containing a varying number of ejaculates to yield complete genetic profiles. In addition, DNA from relatives living within the same household was detected in most cuttings from the children's underwear. In another study, Brayley-Morris et al. (2015) demonstrated that complete DNA profiles could be obtained from laundered semen stains on clothing, even with an 8-month lag time between semen deposition and laundering. All of these studies show that biological material, such as semen, can be transferred from semen-stained clothing to pristine clothing during laundering. However, such data could be interpreted as evidence of sexual assault or activity, especially when found on a child's undergarments or on a complainant's clothing who is unable to articulate the details of a specific event. Such information and scenarios should help forensic biologists interpret evidence from such difficult intrafamilial cases and allow analysts to consider whether the biological material or DNA was directly transferred onto the clothing in question or merely transferred by innocent means.

Touch DNA and wrongful convictions

Improvements in the efficiency and accuracy of DNA typing methods, as well as the increase in the sensitivity of the instrumentation, have created new sources of testable DNA evidence, including touch DNA. As stated previously, touch DNA refers to the genetic information recovered from epithelial (or skin) cells left behind when a person makes contact with another person or another object. During the commission of a crime, an assailant can leave touch DNA behind when s/he has applied force or pressure, which deposits cells on a victim's clothing or other items implicated in the crime. Touch DNA testing uses the same PCR technology and STR analysis to test epithelial cells for DNA as samples from more traditional sources of DNA. The main difference between traditional DNA testing of bodily fluids and touch DNA is the material from which the DNA is collected and extracted, not the method by which the DNA sample is analyzed. But due to these series of technological advances, it is now possible to detect DNA at levels never imagined when DNA fingerprinting started in 1985. In essence, a crime scene investigator only needs to collect a sample with a few epithelial cells to generate a full DNA profile. However, this heightened level of sensitivity can create

misleading results, such as false positives, or reveal the profile of an individual who was not present at the crime scene.

A study by Bolivar et al. (2016) showed that fingerprint brushes used at a crime scene could transfer DNA from one crime scene to another. This study also indicated that the detection of secondary transfer of DNA is enhanced using low-copy number STR typing procedures. This is another example where the forensic analyst will need to carefully consider the various scenarios as to the origin or deposition of biological material or DNA on evidentiary samples.

Notable cases involving touch DNA and secondary transfer of biological evidence have inspired TV and movie producers to create murder mysteries based on this phenomenon of DNA transfer. *Stillwater*, a new (2021) American crime drama film based on the conviction of Amanda Knox, an American college student convicted in Italy for the murder of her roommate, is such an example. In 2007, Knox, her boyfriend, and a third defendant were accused of murdering her roommate. Police suspected that Knox and her boyfriend held the knife against the victim while the third defendant assaulted her. To support this theory, police pointed to a trace amount of the boyfriend's DNA that was found on the victim's bra strap and that both Knox's and the victim's DNA were found on a knife recovered from the boyfriend's apartment. However, the knife did not match the victim's wounds. The Italian Supreme Court acquitted Knox in 2015, partly due to evidence contamination and a misunderstanding of how DNA transfers. This heightened sensitivity of cross-contamination and biological transfer can easily create false positives. For example, analysts are picking up DNA transferred from one person to another by way of an object that both individuals have touched, or from one piece of evidence to another by crime scene investigators or laboratory technicians, or when two items touch each other in an evidence bag. Knox spent 4 years in an Italian prison because of her wrongful conviction and another 3 years trying to clear her name. In 2015, the Italian Supreme Court's report established that the third defendant was the crime's only perpetrator. Although juries and judges have overcome their skepticism in certain cases, acquittals often come months or years after the defendant's life was interrupted.

A new awareness

Since the introduction of DNA profiling, the analysis of DNA from biological evidence has become a daily occurrence in forensic laboratories worldwide. Over the years, the technology to isolate and analyze DNA has become very sophisticated and sensitive. Since DNA testing is sensitive enough to generate full profiles from every cell, great care in the handling of evidentiary samples is critical. In fact, crime scene investigators and forensic analysts in the laboratory handling evidence are required to adhere to stringent and well-established laboratory protocols to minimize and/or avoid contaminating the evidence.

Direct and indirect transfer of biological evidence has been shown to occur between individuals and inanimate objects. And because of direct and indirect transfer, biological material containing DNA can be isolated from a piece of evidence and the individual identified. However, as explored earlier, certain situations could lead to analysts' misinterpretation of the deposition of the DNA. Such "misplaced" DNA evidence and misinterpretations would, no doubt, lead to the wrongful conviction of an innocent person.

Improvements in the efficacy and accuracy of DNA typing methods have provided a resource of evidence for DNA testing. Touch DNA is an example whereby a person can make contact with another person or object during the commission of a crime and leave trace amounts of skin cells or DNA. Such analyses of touch DNA have led to uncountable convictions; however, this heightened level of efficacy and sensitivity can create misleading results or reveal a profile of an individual that was not present at the crime scene or involved in the crime. Needless to say, as science changes and technology consistently improves, both the legal and scientific communities will need to work together in the interest of justice to fully appreciate the magnitude of such findings when DNA is detected from multiple sources.

References

Anker, P., Lyautey, J., Lederrey, C., & Stroun, M. (2001). Circulating nucleic acids in plasma or serum. *Clinica Chimica Acta, 313*, 1453−1456.

Bolivar, P., Tracey, M., & McCord, B. (2016). Assessing the risk of secondary transfer via fingerprint brush contamination using enhanced sensitivity DNA analysis methods. *Journal of Forensic Sciences, 61*(1), 204−211.

Botezatu, I., Serdyuk, O., Potapova, G., Shelepov, V., Alechina, R., Molyaka, Y., Anan'ev, V., Bazin, I., Garin, A., Narimanov, M., Knysh, V., Melkonyan, H., Umansky, S., & Lichtenstein, A. (2000). Genetic analysis of DNA excreted in urine: A new approach for detecting specific genomic DNA sequences from cells dying in an organism. *Clinical Chemistry, 46*(8), 1078−1084.

Brayley-Morris, H., Sorrell, A., Revoir, A. P., Meakin, G. E., Court, D. S., & Morgan, R. M. (2015). Persistence of DNA from laundered semen stains: Implications for child sex trafficking cases. *Forensic Science International: Genetics, 19*, 165−171.

Daly, D. J., Murphy, C., & McDermott, S. D. (2012). The transfer of DNA from hands to glass, fabric and wood. *Forensic Science International: Genetics, 6*, 41−46.

Edwards, A. A., Civitello, H. A., Hammond, H. A., & Caskey, C. T. (1991). DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. *The American Journal of Human Genetics, 49*, 746−756.

Edwards, A. A., Hammond, H. A., Jin, L., Caskey, C. T., & Chakraborty, R. (1992). Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. *Genomics, 12*, 242−253.

Fonnelop, A. E., Egeland, R., & Gill, P. (2015). Secondary and subsequent DNA transfer during criminal investigation. *Forensic Science International: Genetics, 17*, 155−162.

Garcia-Olmo, D. C., Ruiz-Piqueras, R., & Garcia-Olmo, D. (2004). Circulating nucleic acids in plasma and serum (CNAPS) and its relation to stem cells and cancer metastasis: State of the issue. *Histology & Histopathology, 19*, 575−583.

Goray, M., Mitchell, R, R. J., & Oorschot, A. H. V. (2010). Investigation of secondary DNA transfer of skin cells under controlled tests conditions. *Legal Medicine, 12*(3), 117−120.

Goray, M., Mitchell, J. R., & van Oorschot, R. A. H. (2012). Evaluation of multiple transfer of DNA using mock case scenarios. *Legal Medicine, 14*, 40−46.

Ladd, C., Adamowicz, M. S., Bourke, M. T., Scherczinger, C. A., & Lee, H. C. (1999). A systematic analysis of secondary DNA transfer. *Journal of Forensic Sciences, 44*(6), 1270–1272.
Lehmann, V. J., Mitchell, R. J., Ballantyne, K. N., & van Oorschot, R. (2013). Following the transfer of DNA: How far can it go? *Forensic Science International: Genetics, 4*(1), e53–e54.
Lowe, A., Murray, C., Whitaker, J., Tully, G., & Gill, P. (2002). The propensity of individuals to deposit DNA and secondary transfer of low level DNA from individuals to inert surfaces. *Forensic Science International, 129*, 25–34.
Milestone, L. M. (2004). Epidermal desquamation. *Journal of Dermatological Science, 36*(3), 131–140.
Noel, S., Lagace, K., Rogic, A., Granger, D., Bourgoin, S., Jolicoeur, C., & Seguin. (2016). DNA transfer during laundering may yield complete genetic profiles. *Forensic Science International: Genetics, 23*, 240–247.
Rudin, N., & Inman, K. (2002). *An introduction to forensic DNA analysis* (2 ed.). Boca Raton, FL: CRC Press.
Sender, R., Fuchs, S., & Milo, R. (2016). Are we really vastly outnumbered? Revisited estimates for the number of human and bacteria cells in the body. *Cell, 164*, 337–340.
Smith, P. A. (2016). *When DNA implicates the innocent*. http://www.scientificamerican.com/article/when-dna-implicates-the-innocent/.
Van Oorschot, R. A. H., & Jones, M. K. (1997). DNA fingerprints from fingerprints. *Nature, 387*, 767.
Warne, D. C., Watkins, C., Bodfish, P., Nyberg, K., & Spurr, N. K. (1991). Tetranucleotide repeat polymorphism at the human B-actin related pseudogene 2 (actbp2) detected using the polymerase chain reaction. *Nucleic Acids Research, 19*, 6980.

Further reading

Kafarowski, E., Lyon, A. M., & Sloan, M. M. (1996). The retention and transfer of spermatozoa in clothing by machine washing. *Canadian Society of Forensic Science Journal, 29*(1), 7–11.
National Institute of Justice and Centers for Disease Control and Prevention. (1998). *Prevalence, incidence and consequences of violence against women survey*. https://www.rainn.org/statistics/victims-sexual-violence.

CHAPTER 3

Overview of NGS platforms and technological advancements for forensic applications

Anna Barbaro[1,2]

[1]Department of Forensic Genetics, Studio Indagini Mediche E Forensi (SIMEF), Reggio Calabria, Italy; [2]Worldwide Association of Women Forensic Experts (WAWFE), Rome, Italy

Introduction

Since its first introduction in 1985, DNA typing has evolved rapidly and has become the most powerful technique in criminal investigations. DNA profiles obtained from an evidence can be compared to the DNA profiles of a suspect, a victim, or other samples, or they can be searched in national DNA databases. This allows for linking a suspect with the victim, establishing the presence of an individual at a scene, and also connecting different crimes (An et al., 2012). Moreover, DNA has been proven useful in parentage testing, establishing immigration eligibility, and genealogical studies.

Routine forensic DNA tests consist of the analysis of length polymorphisms (short tandem repeats, or STR) found either autosomal or in sex chromosomes (X−Y). Besides, mitochondrial DNA (mtDNA) analysis is also carried out to study the maternal lineage of an individual in sample-limiting conditions.

These sequences differ on the basis of the total length of the repeated area, the length of the repeated units, and the number of units. Routinely, STR markers are amplified by polymerase chain reaction (PCR) using sequence-specific primers, and then PCR products are separated and detected by capillary electrophoresis (CE).

This method unfortunately has some limitations, such as the inability to analyze multiple genetic polymorphisms in a single reaction using a single workflow, partial loss of information from degraded or low-quality DNA samples, and difficulties in defining components in a mixture (Yang et al., 2014).

CE is also the basis of the Sanger method, which represented the gold standard for DNA sequencing for approximately 40 years. For it, Frederick Sanger was awarded the Nobel Prize in chemistry in 1980. The Sanger method, also known as first-generation sequencing, is based on the selective incorporation by DNA polymerase of chain-terminating dideoxynucleotide (ddNTp) during PCR (Sanger & Coulson, 1975, Sanger et al., 1977).

Next Generation Sequencing (NGS) Technology in DNA Analysis
ISBN 978-0-323-99144-5, https://doi.org/10.1016/B978-0-323-99144-5.00019-6
© 2024 Elsevier Inc.
All rights reserved.

Initially, radioactively labeled ddNTPS were used for chain termination, which was subsequently replaced with fluorescently labeled nucleotides with the introduction of automatic sequencers.

CE-based Sanger sequencing is generally used in forensics to analyze specific regions of mtDNA.

Sanger sequencing is capable of producing a template sequence per reaction, and because of this, it is the best choice for analyzing small numbers of gene targets and samples.

Anyway, the method has some limitations, such as poor quality in the first 15–40 bases of the sequence due to primer binding, deteriorating quality of sequencing reads after 700–900 bases, and reduced extension of the sequence that is generally limited to the regions of interest.

These limitations induced forensic scientists worldwide to explore new technologies, and in the last few years, the introduction of next-generation sequencing (NGS) has greatly improved forensic DNA analysis.

Next-generation sequencing platforms

In the last decades, innovative technologies have been developed and have found application in several fields of biological research (Bruijns et al., 2018).

NGS was introduced for commercial use in 2005 (Margulies et al., 2005). It was initially called "massively parallel sequencing", because it enables simultaneous sequencing of many DNA strands in lieu of one strand at a time as in Sanger sequencing.

NGS technology is used to determine DNA or RNA sequences in entire genomes or targeted regions. It offers high-throughput capacity, reduced costs, and high speed, allowing it to perform a wide variety of applications. Because of this, it has revolutionized the biological sciences, and it has been applied in various fields such as ancient DNA analysis (Poinar et al., 2006), agrigenomics (Goddard & Hayes, 2009), disease diagnosis (McCarth et al., 2013), and forensics (Weber-Lehmann et al., 2014).

NGS enables performing millions to billions of sequencing reactions simultaneously in multiple samples, as well as analyzing different types of genomic traits in a single run. Additional advantages of NGS include lower sample input requirements, higher accuracy, and the ability to detect variants at lower allele frequencies than with Sanger sequencing.

Using CE Sanger sequencing, the Human Genome Project took over 10 years with a cost of around $3 billion. NGS now allows the sequencing of the whole human genome within 3 days at a cost of around US $1000.

Depending on the length of sequences produced second- and third-generation sequencing methods are available, and they are defined as short-read and real-time long-read technologies, respectively (Liu et al., 2012; Slatko et al., 2018).

Second-generation sequencing

The NGS workflow basically includes three steps: sample preparation, nucleic acid sequencing, and data analysis (Kumar et al., 2019).

There are several different strategies used to produce clonal template populations: emulsion, solid-state, and DNA nanoball generation. In all cases, the first step consists of the fragmentation of the DNA template.

In emulsion PCR, DNA templates are fragmented and ligated to adapter sequences, and then they are captured inside micelles (aqueous droplets) using a bead covered with complementary adapters, primers, dNTPs, and DNA polymerase. The PCR reaction is carried out within the micelle, and each bead becomes covered with thousands of copies of the target DNA sequence. This strategy is used by 454 (Roche), SOLiD (Thermo Fisher), GeneReader (Qiagen), and Ion Torrent (Thermo Fisher).

In solid-phase bridge amplification, after fragmentation, DNA templates are ligated to adapters, and then they bind with primers immobilized onto a solid support. The free extremity interacts with other primers, creating a bridge structure. Using PCR, replicates stands are produced from the immobilized primers while unbound DNA is removed.

In solid-phase template walking, fragmented DNA templates are ligated to adapter sequences, and then they bind with primers immobilized onto a solid support. Through the PCR process, a second strand is produced. The new double-strand is partially denatured, allowing the free end of the original template to bind with another primer sequence. Reverse primers are used to initiate strand displacement to generate additional free templates, each of which can bind to a new primer.

This strategy is used by **Illumina platforms**.

In DNA nanoball generation, the DNA template is fragmented and ligated to the first of four adapters. The template is amplified, circularized, and cleaved. This process is repeated for all adapters. The final product is a circular template including four adapters, each separated by a template sequence. In this way, a large mass of concatamers (nanoball DNA, or nDNA) is generated.

This strategy is used by **Complete Genomics (BGI)**.

Sequencing methods are grouped into two major categories: Sequencing by hybridization (SBH) and sequencing by synthesis (SBS). The SBH method was developed in the 1980s. It relies on the use of DNA oligonucleotides arrayed on filters, which are then hybridized into labeled fragments of the DNA to be sequenced. The method is based on the use of specific probes to interrogate specific sequences and allows the analysis of large sequences by overlapping information. It is generally used for diagnostic applications such as the identification of single nucleotides polymorphisms (SNPs) related to diseases and the identification of chromosome rearrangements and abnormalities (Slatko et al., 2018; Church, 2006; Drmanac et al., 2002; Hanna et al., 2000; Qin et al., 2012). The SBS method represents a further development of Sanger sequencing. It involves repeated cycles of synthesis to incorporate additional nucleotides into the

growing chain. SBS approaches include cyclic reversible termination (CRT) or single-nucleotide addition (SNA). CRT uses terminator molecules (3′-OH group blocked) for preventing elongation. In SNA each of the four nucleotides is added iteratively to ensure only one dNTP is responsible for the signal. Thus, the absence of the next nucleotide in the reaction prevents elongation (Goodwin et al., 2016). Current SBS methods differ from the original Sanger sequencing because they rely on shorter reads with high sequence coverage ("massively parallel sequencing") of millions to billions of short DNA sequence reads (50–300 nucleotides). The most commonly diffused next-generation technologies are based on SBS approaches, such as the 454, Ion Torrent, and Illumina platforms.

The first NGS method developed was Roche 454 pyrosequencing commercialized in 2007. Pyrosequencing is a real-time sequencing alternative to the Sanger method (Ronaghi et al., 1996). It is based on the detection of pyrophosphate release and the consequent generation of light after nucleotide incorporation. As a dNTP is incorporated into a strand, an enzymatic cascade occurs, producing a bioluminescence signal that is detected by a charge-coupled device camera. By reading optical signals it is possible to identify which base is incorporated. Pyrosequencing is cheaper and faster than Sanger sequencing (Andreasson et al., 2002; Goodwin et al., 2016; Margulies et al., 2005). Using the Genome Sequencer FLX System, the method was able to sequence 400–600 megabases in 10 h.

In 2010, Applied Biosystems began the commercialization of the Ion Torrent, a non-optical sequencing platform with integrated circuits based on the use of a semiconductor ion. This method relies on the iterative addition of each of the four dNTPs to a sequencing reaction to ensure only one nucleotide is responsible for the signal. The absence of the next nucleotide stops the sequencing reaction preventing elongation. During DNA synthesis when a new dNTP is incorporated, a proton is released (H^+), and there is a small change in pH that is detected by a sensor (a field-effect transistor). This permits the identification of the bases (A, C, G, and T) incorporated into the chain. The chemical signals are translated into digital information. Different types of chips are available on the Ion Torrent platforms, enabling complete sequencing in less than 2 h. Currently, two Ion Torrent instruments are available: The Ion GeneStudio S5 System and the Ion Torrent Genexus System.

Illumina is the most widely used sequencing technology because it offers effective platforms capable of guaranteeing correct results, from small, low-throughput benchtop instruments to large, ultrahigh-throughput instruments for whole-genome sequencing. The system uses terminator molecules that are similar to those used in Sanger sequencing. dNTPs are labeled with a base-specific, cleavable fluorophore and blocked by a 3′-O-azidomethyl group. During PCR, all four dNTPs are added, and when a single dNTP is incorporated it reversibly blocks chain elongation. The color emitted by TIRF (total internal reflection fluorescence microscopy using two or four laser channels) allows us to identify which nucleotide was incorporated in each cluster. The dye is then cleaved,

and 3′-OH is regenerated with the reducing agent tris-phosphine. After the removal of the fluorophore and blocking group a new cycle can begin.

The first Illumina platform was launched in 2007, and it was called Genome Analyzer. Initially, it sequenced 6 million amplified fragments per lane but generated only around 30 base reads, but then it was improved to an 80-gigabase output.

HiSeq was the second instrument released by Illumina in 2010, followed by the HiSeq X10, which was able to further increase the number of fragments that could be analyzable. Today, the NextSeq and NovaSeq machines are the most diffused, along with some benchtop sequencers (iSeq 100, MiniSeq).

In 2012, Qiagen bought the Intelligent BioSystems CRT platform, which was commercialized and relaunched in 2015 as the GeneReader 22, which is an all-in-one platform able to perform all steps from sample preparation to analysis. Because of this, the GeneReader system is connected with the QIAcube sample preparation system and the Qiagen Clinical Insight platform for variant analysis.

The GeneReader uses the same approach used by Illumina: dNTPs are blocked by a 3-O-allyl group, and some of the bases are labeled with a base-specific, cleavable fluorophore. Unincorporated bases are washed away, and the sample is imaged by TIRF using four laser channels. The dye is then removed, and the 3′-OH is regenerated by adding a mixture of palladium and $P(C_6H_4SO_3Na)_3$ (TPPTS).

The GenapSys Sequencing Platform is a small, fast, and cheap desktop instrument. It's an alternative NGS system useful for the analysis of small panels due to its affordability and ease of use. The system uses complementary metal oxide semiconductor sequencing chips with millions of sensors (electrodes) on their surface that detect differential electrical changes caused by nucleotide incorporation. It is used for many applications, such as targeted sequencing, small genome sequencing, gene editing validation, RNA sequencing, and targeted single-cell assay sequencing.

Some problems related to second-generation sequencing, such as the difficulty of reliably resolving large genomic rearrangements or the complexity in data assembly, were overcome by the development of third-generation sequencing.

Third-generation sequencing

Third-generation NGS is a group of DNA sequencing methods that were first described in 2009 and are still under active development. The aim of third-generation sequencing is to sequence longer DNA (and RNA) sequences (>1 kb to 2 Mb). The approach is based on single molecule reading in real-time, which allows the generation of long reads faster than with previous sequencing techniques since clonal amplification steps are eliminated (Ozsolak, 2012).

These methods can be grouped into three categories: the first category is where single polymerase molecules are monitored during DNA synthesis (e.g., PacBio); the second category is based on the transport of DNA molecules through a nanopore membrane

(e.g., Oxford Nanopores); and the third category consists of methods that use direct imaging of individual DNA molecules by advanced microscopy (e.g., VisiGen). (Schadt et al., 2010; Stoloff & Wanunu, 2013; Thompson & Milos, 2011).

PacBio sequencing, also known as single-molecule real-time (SMRT) sequencing, enables the analysis of very long fragments (up to 30–50 kb). The SMRT method is based on the binding of the DNA to be sequenced with DNA polymerase at the bottom of a well in an SMRT flow cell, and the DNA strand to be sequenced progresses through it. Imaging occurs only at the bottom of the well when fluorescently labeled dNTPs are incorporated by DNA polymerase. The four nucleotides are labeled with different fluorophores so that each base is identified as it is incorporated into the growing DNA chain. Imaging is related to the rate of nucleotide incorporation. The incorporation simultaneously occurs in parallel in up to 1 million wells on a single chip within the SMRT cells. This technique is fast, and longer reads can be generated than with previous generations of sequencing techniques (Van Dijk et al., 2014).

In 2014, Oxford Nanopore Technologies released the MinION platform based on the transport of DNA through a nanopore membrane embedded in a lipid bilayer or a synthetic polymer (Mikheyev & Tin, 2014). The detection principle of nanopore systems is based on the ion current changing, which is generated when a charged molecule passes through the membrane. When the electric field is applied DNA molecules pass through the pore, and the current of other ions is partly blocked. The decrease in current amplitude is used to determine the nucleotide sequence of the DNA passing through the membrane since each of the four different dNTPs produces distinct current levels (Bruijns et al., 2018; Haque et al., 2013; Schneider & Dekker, 2012). With a nanopore system, a single DNA molecule can be read rapidly without the need for PCR amplification or the use of expensive fluorescent dye labels.

The VisiGen Biotechnology approach is based on fluorescence resonance energy transfer (FRET) and the use of a special DNA polymerase, which acts as a 'real-time sensor' for modified nucleotides. When fluorescently labeled dNTPs are incorporated into a single DNA strain, a FRET occurs. The frequency of the FRET signal varies depending on the label incorporated in the nucleotides, so it is possible to determine base sequences. This approach, not requiring cloning and amplification, reduces the analysis cost and produces ability distinguishing long reads (around 1 kb) in a short time (Pareek et al., 2011) (Table 3.1, Table 3.2).

NGS forensic applications

As technology advances, the NGS method shows huge promise to become a routine method in forensic applications. NGS allows examiners to generate data that spans the human genome in a single, targeted assay. Such a test may be useful in many aspects, as described below.

Table 3.1 Main advantages and disadvantages of NGS second and third generation.

	Advantages	Disadvantages
Second-generation NGS	High accuracy Ability to sequence fragmented DNA Low cost	Short sequencing reads (between 200 and 300 bases) Difficulty in resolving structural variants or distinguishing highly homologous genomic regions Not suitable for sequencing traits containing large numbers of repetitive sequences
Third-generation NGS	Very long sequence reads Ease library preparation Good identification of epigenetic markers Portable technologies	Weak signals from individual fragments Overall lower accuracy

Criminal casework: Using NGS, forensic laboratories can analyze multiple sets of forensic-relevant markers (STR and/or SNPS) in a single assay. This allows for obtaining a wide range of information from degraded, scarce, or mixed DNA samples from cold cases.

Forensic DNA databases: Each year millions of DNA profiles are stored in national DNA databases. NGS can help forensic laboratories to produce high-quality forensic DNA profiles faster than with the CE method.

Disaster victim identification (DVI): In the case of mass disasters, it is relevant to obtain as much information as possible from highly compromised samples in a short time. NGS greatly helps in this regard because of its ability to analyze simultaneously hundreds of markers in a single test.

Missing person identification: In missing person identification cases samples are often compromised. NGS allows sequencing of both nuclear and mtDNA on a single platform, overcoming difficulties encountered with human remains samples.

Short tandem repeats and single nucleotide polymorphism analysis

STR analysis is essential in criminal casework because of the large national DNA databases, which include STR profiles from criminal offenders, evidence, missing persons, and samples from old cases. Consequently, forensic NGS assays must be able to sequence STR loci (Børsting et al., 2014). STR analysis represents the most commonly used method in forensics due to its multiple advantages, such as multiplex amplification, precise allele discrimination, and low DNA template requirements. However, simultaneous detection of a large number of STR markers is difficult due to the technical limitations of

Table 3.2 Summary of common NGS platforms.

Platforms	Technology	Analysis time (average)	Read length (average)	Throughput Mb/h (average)
454/Roche FLX (2008)	Pyrosequencing	4 h	200–300 bp	20–30
Illumina/Solexa (2008)	Reversible dye terminator	2–3 days	30–100 bp	20
ABI/SOLiD (2008)	Oligonucleotide 8-mer chained ligation	8 days	20–45 bp	5–15
Illumina MiSeq (2019)	Reversible dye terminator	1–3 days	2 × 75 –2 × 300 bp	170–250
Oxford Nanopore MinIon (2014)	–	3 days	13–20 kb	700
Illumina NovaSeq (2019)	Reversible dye terminator	1–2 days	2 × 50 –2 × 150 bp	22,000–67,000
Ion Torrent (2019)	Native dNTPs, ion detection	2.5–4 h	200–600 bp	110–920
BGI MGISEQ-T7 (2019)	Nanoball technology	1 day	2 × 150 bp	250,000
Pacific biosciences SMRT (2019)	Phospholinked fluorescent nucleotides	10–20 h	10–30 kb	1300

fluorescent-based CE technology (Yang et al., 2014). NGS technology shows many potential advantages, such as high throughput, the capability for simultaneous detection of a wide range of autosomal and sexual STR loci, and the ability to distinguish alleles with similar lengths.

Routinely, Y-STRs are used to identify the male component of DNA mixtures, especially in sexual assault cases where a high female background is present, or to reconstruct paternal relationships through the paternal lineage. The ability to analyze simultaneously autosomal and sexual STRs greatly increases the analysis efficiency and reduces the cost while allowing better identification, especially in the case of mixed samples. Furthermore, it has been shown that Y chromosome sequencing could help in distinguishing between mixed male samples from the same male lineage (Yang et al., 2014), since in a study by NGS on two men who had the same ancestor 13 generations ago, around 10 million nucleotides of the Y chromosome were sequenced (Xue et al., 2009), and four genetic differences were found.

In order to standardize the nomenclature used for describing STR complex sequences obtained by NGS and to enable comparison with routinely used nomenclature derived from CE systems and with data stored in the national DNA databases, the DNA Commission of the International Society for Forensic Genetics has developed some recommendations (Parson et al., 2016). SNPs are useful in forensics due to their abundance in the human genome, their low mutation rate, and their smaller size than STRs, which increase the chance of success when analyzing degraded samples. Because of this, they are useful in human identification, ancestry studies, and phenotyping prediction (Sobrino & Carracedo, 2005). Furthermore, the high throughput of NGS can assist in targeting SNP sets not routinely available with traditional CE-based methods. In fact, NGS, in comparison with traditional methods (SNaPshot or array systems), allows us to analyze a larger number of markers simultaneously even from low DNA quantities(Seo et al., 2013). In addition, it's possible to target microhaplotypes that have shown promising in forensics for identification and ancestry purposes (Pakstis et al., 2012). NGS technologies have led to the discovery of millions of SNPs, and several custom panels of large SNP assays have been developed by several authors (Wang et al., 2019; Zhang et al., 2017).

Forensic phenotyping and ancestry studies

Forensic phenotyping includes the prediction of externally visible characteristics (e.g., eye color, hair color, and skin color) and biogeographic ancestry, that is, the geographical origin of an individual and the person's age. Tests for the identification of these characters have been developed and validated, along with adequate predictive statistical models. This allows mass screening when there are no suspects or matches with profiles in DNA databases; it's useful in cases of identification of human remains (DVI cases) (Shneider Peter et al., 2019).

In 2011, Walsh from Erasmus University developed the *IrisPlex*, the first assay for accurate prediction of blue and brown eye color, followed in 2013 by the HIrisPlex system for simultaneous prediction of hair and eye color with an accuracy of 90% (Walsh et al., 2011, 2013). In 2018, Walsh developed the HIrisPlex-S system for the analysis of 41 SNPs for the joint prediction of the color of the eyes, hair, and skin. Blue and brown eyes are predicted more accurately than eyes that are neither brown nor blue; red and black hair more accurately than blonde and brown; and dark skin colors more accurately than light (Chaitanya et al., 2018). The prediction of facial shape is one of the main challenges when studying phenotyping to obtain the final "DNA facial composite". Some genetic markers associated with facial features are correlated with craniofacial development and, consequently, linked to normal variation in facial shape; others are related to facial syndromes and deformities and disease studies (such as cleft palate, cleft lip, and other craniofacial dysplasia). However, the human face is a complex feature composed of different characteristics (for example, eye shape and distance between eyes, nose, and mouth shape), whose development involves molecular and environmental interactions that are not fully understood.

In 2009, Klimentidis (Klimentidis & Shriver, 2009) started to study facial features by DNA test and association analysis and validated their results using facial reconstruction molecular photofitting. By joining information about genotype, sex, ancestry, and the effects of particular alleles on facial features, information can be discovered (Claes et al., 2014). In 2020, Liu found 12 SNPs from 10 genes associated with one or more facial shape traits and was able to provide reference data for DNA-based face prediction in the Han Chinese population (Liu et al., 2020). Since forensic DNA phenotyping (FDP) goes beyond standard forensic DNA typing, its legalization is under debate in different countries due to ethical issues (Shneider Peter et al., 2019).

In 2017, the VISible Attributes through GEnomics (VISAGE) Consortium was created as a collaboration between 13 partners from academic, police, and justice institutions belonging to 8 European countries. VISAGE received funds from the European Union's Horizon 2020 with the aim to overcome the general limitation of current forensic DNA analysis by developing new technologies able to provide new tools in addition to conventional DNA analysis, such as predicting the appearance and age of an evidence donor (http://www.visage-h2020.eu/) in order to obtain from DNA a sketch of unknown perpetrators.

An ancestry study is based on the genetic characteristics that a person has inherited from their biological ancestors. The more the distance between the geographical regions of origin of two people, the more the genetic differences between them. These differences are due to mutations, migrations, local selection, isolation, and heredity. Because of this, some DNA markers are very common in particular geographic regions and rare in others. Markers located in the autosomes are inherited from both parents and, therefore, reflect their geographical region of origin. Markers placed on the Y

chromosome are transmitted only from father to son and therefore exclusively reflect the geographical origin of the ancestors through the male lineage. mtDNA markers are transmitted only from mother to child and therefore exclusively reflect the origin of the ancestors (male or female) through the female (maternal) lineage. Because of this ancestry analysis involves the study of all these ancestry informative markers (AIMS) and a prediction based on the maximum likelihood obtained using a classification algorithm (Elaine et al., 2018; Phillips et al., 2007, 2014). For example, in the 2004 Madrid terrorist attack, a suspect was identified by inferring the original population using 34 autosomal SNPs related to the ancestry of the population (Phillips et al., 2009).

Epigenetic analysis

Epigenetics consists of the study of external not heritable factors that may temporally modify genes impacting their expression, i.e., chemical reactions due to exposiure to the environment or as a consequence of a diet or lifestyle. Epigenetic markers have several applications in forensic science, for example, for age prediction, tissue identification, and monozygotic (MZ) twin differentiation. Aging is strongly correlated with changes in DNA methylation (Bocklandt et al., 2011; Frumkin et al., 2011). The molecular estimation of age is based on the fact that the activity of some genes changes during aging. One type of age-dependent gene regulation is the increasing or decreasing degree of cytosine methylation in CpG dinucleotides in the promoter regions of certain genes (Johnson et al., 2012).

In forensic investigations, the ability to predict the biological age of the donor of a sample can provide relevant information in cases where no suspects or matches in the DNA database are found (Freire-Aradas et al., 2017). From a clinical point of view, it's useful to estimate the life expectancy of an individual (Horvath, 2013; Horvath & Raj, 2018). DNA methylation analysis available to date allows for estimating age with an average error of $\pm 4-5$ years (Freire-Aradas et al., 2020; Álvarez-Dios et al., 2020). Epigenetic markers have been found also to be useful to distinguish the type of biological material found at crime scenes. In fact, in some cases, traditional serological methods used for identifying body fluids are not enough due to low sensitivity (especially in the case of trace samples) and the inability to distinguish, for example, menstrual blood from regular blood (Titia Sijen & Harbison, 2021). An alternative approach is to utilize patterns of epigenetic DNA methylation (Lee, 2015; Park, 2014), since genome-wide methylation allows the identification of variants on the methylation degree of CpG sites, useful to distinguish different tissues (Forat et al., 2016; Park et al., 2014).

Epigenetic approaches based on NGS technology include whole-genome bisulfite sequencing, methylation bead chips, and immune-precipitation sequencing (Xu et al., 2012). Since it's crucial to be able to identify trace samples, it has been found that low DNA amounts (100 pg) are successfully analyzed through genome-wide amplification of a bisulfite-modified DNA template, followed by quantitative methylation detection

using pyrosequencing (Paliwal et al., 2010; Xu et al., 2012). The ability to differentiate MZ twins is another relevant topic in forensics useful for the correct identification of the author of a crime or a victim. Since twins have the same DNA, the analysis of autosomal and sexual STRs, SNPs, or mtDNA cannot differentiate them. Methylation studies have been found useful to solve this issue: For 65% of the pairs the twins show almost identical epigenetic profiles, but for the remaining 35% there are significant differences. In 2010, the Beijing Genomics Institute and TwinsUK, the twin research group based at King's College London, started a research project to study by NGS technology in 5000 pairs of twins the methylation patterns of 20 million sites looking for differences in epigenetic signals useful to explain why many twins don't develop the same features or diseases (Yang et al., 2011). In 2013, 92 significantly methylated CpG sites were found, representing the potential targets for distinguishing between MZ twins (Li et al., 2013). Similarly, in 2014, extremely rare mutations were analyzed to differentiate between MZ twins by ultradeep NGS (Weber-Lehmann et al., 2014).

RNA analysis

DNA analysis allows the identification of an individual but does not give contextual information about the sample. In the last few years, RNA has been found useful for several applications. Several studies have been carried out to understand if gene expression can be useful for solving this issue. RNA types such as mRNAs and microRNAs have been shown to be tissue-specific, and they can provide high specificity for species identification and body fluid identification. Markers have been identified for the forensically most relevant body fluids (i.e., blood, saliva, semen, vaginal secretion, and menstrual blood). Additionally, it has been found that the mechanism of RNA degradation can give important information about sample deposition time (Ingold et al., 2018; Simard et al., 2012; Zubakov et al., 2008) and the determination of postmortem interval (Courts & Madea, 2010). Although RNA is highly susceptible to degradation by environmental RNAses, the use of NGS technology can greatly improve the analysis because millions of RNA sequences can be rapidly analyzed (Hanson et al., 2018).

Microbial forensics

For over 100 years, microbiology played a minor role in forensic science (MacCallum & Hastings, 1899), and only in the early 2000s, with the rise of bioterrorism was so-called "microbial forensics" born.

Microbial forensics is a discipline developed by the Federal Bureau of Investigation after the Anthrax attack in the USA on September 18, 2001, due to the serious consequences produced by microbiological terrorist attacks. It consists of the fast and accurate detection and identification of microorganisms with the aim of tracing the source of the microbes (Beecher, 2006; McEwen et al., 2006).

Microorganisms are highly different and ubiquitous, and they can be found everywhere. They are abundant in the human body and inside it, in the environment, and on objects (Oliveira & Amorim, 2018; Sender et al., 2016; Vázquez-Baeza et al., 2018). Because of this, microbial profiles could be used to complement traditional criminal investigations (Jake et al., 2021) for establishing sample provenance, helping identify individuals, determining the cause of a death, and estimating PMI (Jake et al., 2021; Phan et al., 2020; Schmedes et al., 2017). Since NGS has the advantages of high throughput, multiplexing capability, and accuracy, it is suitable for rapid whole-genome typing of microorganisms in forensic or epidemiological investigations, allowing the identification of different bacterial taxa and strains (Brenig et al., 2010; Fierer et al., 2010; Giampaoli et al., 2014; Lilje et al., 2013; Wooley et al., .2010).

Mitochondrial DNA sequencing

For years sequencing of mtDNA has been used in forensic casework especially for the analysis of scarce or degraded samples (i.e., teeth, ancient bones, and hairs without roots) when STR analysis does not provide a full STR profile (Børsting & Morling, 2015; Parson, 2004, 2007). Mt-DNA has a higher copy number per cell than nuclear DNA, so it can survive environments where nuclear DNA cannot. Because of this, it can be a powerful tool for human identification. NGS provides deep coverage to obtain genetic data from small forensic mtDNA samples in not ideal conditions. Other advantages consist of high sensitivity, increased heteroplasmy detection, and a feasible method (Eduardoff et al., 2017). Different PCR strategies exist for amplifying the mitochondrial genome: two long segments for whole genome amplification of good-quality DNA, multiple small fragments (~2 kb) for whole genome amplification of compromised samples, and many overlapping small amplicons for amplification of either the control region or entire genome of degraded samples (Lyons et al., 2013; McElhoe et al., 2014; Parson et al., 2015). Multiplexes are available commercially for amplifying either the control region or the entire mitochondrial genome in 100−400 bp amplicons (Thermo Fisher Scientific, Promega).

The ability to analyze the entire mitochondrial genome significantly increases the discriminatory power in comparison with control region analysis alone, allowing for maximal resolution of matrilineal geographic ancestry. Additionally, for severely degraded samples, probe capture-NGS methods have been shown to provide results with old or extensively damaged mtDNA (King et al., 2014; Peck et al., 2016; Templeton et al., 2013).

Animal and plant DNA analysis

Species determination represents a relevant tool in criminal investigations, for example, in illegal animal traffic, in the food industry, or in mass disasters to distinguish human remains from animal remains (Yang et al., 2014). Routinely, species identification is

performed using species-specific PCR primers. However, in many cases, no a priori species information is available. In these cases, NGS technology allows identification by sequencing the entire genome and comparing it with a species' reference genome. In addition, NGS is useful in evolutionary studies to determine the genome of extinct organisms (e.g., Mammut), in veterinary for the identification of genes responsible for the development of diseases or for disease resistance, and in studies about animal breeding (Dunisławska et al., 2017).

In 2012, the 1000 Bull Genomes Project was created with the aim of comparing whole-genome sequences of cattle by NGS. It includes 40 international partners, 2700 dairy and beef animals, and has identified close to 90 million genetic variants (Hayes & Daetwyler, 2019). In 2018, the US Department of Agriculture-National Animal Health Laboratory Network started a project aimed at evaluating new technologies and developing standardized methods for improving studies about animal diseases (Harris et al., 2021). In addition, NGS may have a great impact in agrigenomics because it allows to perform rapidly and in a cost-effective manner plant whole genome sequencing, exome or transcriptome sequencing (Fullwood et al., 2009), so to discover genetic markers involved in plant development, tolerance to stress, or medicinal plant breeding, and also for the discovery of domestication genes in crop plants and their wild relatives (Henry, 2012; Sing et al., 2015). Using NGS the genome sequences of several plant species have been completed, and some plants and animals in traditional Chinese medicines have also been studied (Yang et al., 2014). Due to their great abundance and variability, SNPs are currently the marker of choice in plant genetic research and breeding. The first plant sequenced was *Arabidopsis thaliana,* followed by rice (Kaul et al., 2000; Yu et al., 2002) (Table 3.3).

Forensic NGS commercial systems

In 2015, a review article was published in which NGS technology was called "the future of forensic DNA analysis" (Børsting & Morling, 2015). The use of NGS in forensic genetics has been debated, but in the last few years, specific applications for human

Table 3.3 Main NGS applications in forensics (alphabetical order).

Main NGS applications in forensics
Animal and plant DNA analysis
Epigenetic analysis
Forensic phenotyping and ancestry studies
Microbial forensics
Mitochondrial DNA (Mt-DNA) sequencing
RNA analysis
Short tandem repeats (STRs) and single Nucleotides polymorphisms (SNPs) analysis

identification, determination of phenotypic traits, and mtDNA analysis have been developed and tested in many forensic and research laboratories (Alvarez-Cubero et al., 2017; Ballard et al., 2020; King et al., 2014). Initially, the introduction of forensics was slow, mainly due to a lack of accredited sequencers and kits and relatively high error rates in comparison with Sanger sequencing. Currently, a majority of the problematic issues have been solved (Ballard et al., 2020). The technology of NGS was initially mainly used for the study of SNP markers and mt-DNA, but nowadays also autosomal STR markers are analyzed.

In 2014, Thermo Fisher Scientific launched two SNP typing assays designed for the Ion PGM™ System: the HID-Ion AmpliSeq™ Identity Panel for human identification that amplifies 124 autosomal SNPs, including most of the SNPforID and individual identification SNPs (IISNPs), and 34 Y-chromosome SNPs, and the HID-Ion AmpliSeq™ Ancestry Panel (Xavier et al., 2020) for ancestry estimation that includes most of the AIMs (Eduardoff et al., 2015; Guo et al., 2016; Nievergel et al., 2013; Sanchez et al., 2006). Thermo Fisher Scientific aims to develop assays to be used as a supplement to PCR-CE typing and to be effective with scarce samples. With the improvement of sequencing technology, the HID Ion GeneStudio S5 Series sequencers have been launched that allow using a little DNA quantity (as 100 pg) to obtain accurate and reproducible results. Currently, the Precision ID GlobalFiler NGS STR Panel v2 includes the same CE-STR markers available in the GlobalFiler PCR Amplification Kit, and it combines high compatibility with database core loci and superior discrimination power.

The Precision ID Ancestry Panel includes 165 autosomal markers (SNPs) that can provide biogeographic ancestry; the Precision ID Identity Panel uses 34 upper Y-Clade SNPs and 90 autosomal SNPs for enabling DNA typing from degraded or challenging forensic samples. The Ion AmpliSeq PhenoTrivium Panel contains DNA markers for phenotyping, ancestry, and male lineage, for a total of 200 autosomal SNPs and 120 Y chromosomal SNPs in a single assay. The Ion Ampliseq DNA Phenotyping Panel is used to predict hair and eye color and includes 24 SNPs from the HIrisPlex system developed by Susan Walsh from Erasmus University (Walsh et al., 2013).

The Visage Consortium developed a Basic Tool for Appearance and Ancestry prediction from DNA that represents the first FDP laboratory tool that combines DNA analysis of the eye, hair, and skin. It allows the analysis of 153 SNPs into a single multiplex reaction using the AmpliSeq design pipeline and is applied for massively parallel sequencing with the Ion S5 platform. They included markers from IrisPlex, HIrisPlex, and HIrisPlex-S systems (Chaitanya et al., 2018; Walsh et al., 2011, 2013).

For biogeographical ancestry (BGA), the most AIM-SNPs were taken from the EUROFORGEN Global AIMs MPS ancestry panel (Eduardoff et al., 2017; Phillips et al., 2014), supplemented with two additional AIMs SNPs (rs10497191, rs6990312) from the 55 AIMs panel developed at Kidd laboratory (Kidd et al., 2014a, 2014b). Furthermore, 11 additional AIM-SNPs included in the Thermo Fisher Precision ID ancestry panel

were added. The panel has high sensitivity and is able to process low or degraded casework samples, obtaining full SNP profiles with DNA amounts down to 100 pg.

In 2017, Illumina and Telegraph Hill Partners established Verogen, Inc., which is the only global organization dedicated to the development and supply of NGS-based human identification products. The MiSeq FGx Sequencing System represents the first NGS instrument developed and validated for forensic use. It is a platform able to prepare and sequence libraries and analyze data in a single workflow designed for a wide range of applications, including FDP and forensic genetic genealogy. The MiSeq FGx Forensic Genomics System is in use in operational crime laboratories, private service labs, and forensic research institutes worldwide. The ForenSeq Kintelligence Kit has been developed for forensic purposes, and it enables the analysis of 10,230 forensically relevant markers in 8 h with less than 2 h of hands-on time.

The ForenSeq MainstAY Kit simultaneously analyzes 27 autosomal STRs and 25 Y-STR markers with as little as 100 pg, and it's compatible with all existing STR databases. The ForenSeq DNA Signature Prep Kit amplifies 27 autosomal STRs, 8 X-STRs, 25 Y-STRs, 95 autosomal human identification SNPs, 56 autosomal BGA-informative AIMs, and 24 autosomal SNPs associated with phenotypic traits in a single reaction. The multiplex includes, among others, all of the STR loci in the CODIS and European standard sets, most of the SNP *for* ID, and all of the HIrisPlex loci. With over 200 markers in a single multiplex, it eliminates the need to run multiple STR tests, reducing time and costs.

It is the only NGS chemistry approved for uploading STR profiles to the National DNA Index System in the USA. The ForenSeq mtDNA Whole Genome Kit and mtDNA Control Kit allow successful sequencing of the whole genome and the control region of mtDNA inforensic samples even if compromised with minimal DNA input. The forensic laboratory for DNA analysis at the University of Leiden represents the first laboratory in the world accredited to use NGS technology in forensic DNA analysis. A survey conducted in 2017 among 33 European forensic laboratories showed that at least 52% of them had already bought one or more NGS instruments: The MiSeq R/NextSeq (Illumine) and the Ion Torrent PGM/S5 (Thermo Fisher) are the most diffused.

NGS advantages and challenges

NGS using a whole genome or targeted approachs allow simultaneous analysis of a large number of markers such, as STRs and SNPs in parallel with targeted mRNA and small RNA analysis. This makes NGS a very powerful tool in forensic laboratories (Ballard et al., 2020). Traditional STR analysis by CE allows us to determine the repeat number of the STR on the basis of the PCR amplicon size. NGS analysis can determine the full sequence of the PCR product, including the STR repeat region and the flanking areas, providing a precise description of the repeat allele structure and all variants in the

surrounding area. Since repeat structures can be complex (due to the presence of insertions, deletions, substitutions, rearrangements, or SNPs in the flanking regions), this type of approach results in increased allelic variability for many STRs (Ballard et al., 2020; Devesse et al., 2017; Gettings et al., 2015), producing an improvement in the markers discrimination power.

In the case of DNA mixtures, the improvement in marker discrimination is a great benefit for identifying extra sequence-specific alleles that may be masked in the case of identical CE amplicon size. In addition, the increased discrimination reduces the risk of an adventitious match between the alleles of the alleged donor and the alleles of the mixed sample. In this way, it becomes easier to distinguish different DNA components in a complex mixture. The utility of STR flanking variants for mixture analysis has been reported in studies about deletion/insertion polymorphisms present in the flanking regions of STRs (DIP-STRs) (Phillips et al., 2007). Furthermore, it has been found that by NGS analysis, mixture proportions are accurately reflected by read numbers (Devesse et al., 2017).

One notable advantage of NGS analysis in comparison with CE is that separation by amplicon size is not required for multiplex STR assay design. STRs can be amplified using the smallest amplicon length, which improves analysis success in cases of compromised samples. Furthermore, the number of markers that may be amplified simultaneously is not constrained by the fluorescent-dye detection systems that allow coamplification of large panels of markers. In this way, amplification is not limited to STR loci, and SNP markers can also be analyzed in combination with STRs.

The advantages of analyzing SNPs for forensic samples with NGS in comparison with SNaPshot or SNP array systems consist in the fact that it is possible to analyze large numbers of markers simultaneously from low quantities of DNA (Daniel et al., 2015; Grandell et al., 2016; Kidd, et al., 2014a, 2014b; Pakstis et al., 2012). It is also possible to detect sets of SNPs located in very close proximity to each other on a chromosome (microhaplotype) that have been shown to be useful in forensics for both identification and ancestry purposes (Wendt et al., 2017). Traditionally, simultaneous analysis of multiple nucleic acids (DNA and RNA) or different DNA markers (STR and SNPs) has not been utilized in forensics.

When the sample quantity is enough the protocol generally includes consecutive RNA and DNA extractions or alternatively divides the samples for separate isolations (Ballard et al., 2020), and then different analyses are performed. NGS protocols enable simultaneous analysis of DNA and RNA, which in the last decade has been shown to also have forensic relevance, for example, in the case of tissue and body fluid identification, determination of time of death, etc.

The main challenge for the application of NGS in forensics consists of the quantity of input samples required for the analysis because forensic samples are notoriously compromised. Initially, large amounts of DNA were required since the reduced concentration of

nucleic acids greatly influenced the quality of the library preparation and sequencing choice (Cornelis et al., 2017). Nowadays, recent improvements allow the successful utilization of low DNA input (Ballard et al., 2020). With the progressive reduction of the DNA amount required for preparing NGS libraries consequently a wide range of forensic samples could be analyzed to obtain large pieces of information. Another challenge is that, depending on the NGS platform, some factors are variable (e.g., read lengths, system flexibility, runtime). Considerable improvements in read length have been achieved in the pyrosequencing and SBS platforms. Another relevant variable is the overall sequencing capacity, or the total number of clusters (or reads) that are sequenced per run.

A high degree of flexibility in the experimental design is possible on the SBS and semiconductor platforms, which makes them suitable for experiments where the number of samples or the sizes of the analyzed regions are variable from one experiment to another. Finally, in pyrosequencing and the semiconductor platforms the runtime is shorter than in the SBS and sequencing by ligation platforms because in the first one, signal detection is performed in real-time, while in the latter, the least is done by imaging. Finally, the differences among NGS platforms in terms of data format, length of reads, etc. result in the need for diversity in bioinformatics tools to be used for sequence quality scoring, alignment, assembly, and data processing (Wold and Myers, 2008; Yang et al., 2009; Pop and Salzberg, 2008).

Generally, in an NGS run tens or hundreds of short reads are produced, which result in terabytes of raw data. Because of the complexity of data analysis, a variety of software for NGS data analysis has been developed, with several tools available online. Their functions consist mainly of alignment of reads to a reference sequence; de novo assembly; reference-based assembly; base-calling; and genetic variation detection (SNV, Indel). Obviously, experience is essential to the successful analysis of data obtained with NGS technology.

References

Alvarez-Cubero, M. J., Saiz, M., Martínez-García, B., Sayalero, S. M., Entrala, C., Lorente, J. A., & Martinez-Gonzalez, L. J. (2017). Next generation sequencing: An application in forensic sciences? *Annals of Human Biology, 44*(7), 581–592.

Álvarez-Dios, J., Ballard, D., Court, D. S., Branicki, W., Carracedo, Á., & Lareu, M. V. (2020). A comparison of forensic age prediction models using data from four DNA methylation technologies. *Frontiers in Genetics, 11*, 932.

Andreasson H., Asp A., Alderborn A., Gyllensten U., Allen M. (2002) Mitochondrial sequence analysis for forensic identification using pyrosequencing technology, Biotechniques 32 (2002) 124–126, 128, 130–133.

An, J. H., Shin, K. J., Yang, W. I., & Lee, H. Y. (2012). Body fluid identification in forensics. *BMB Reports, 45*(10), 545–553. https://doi.org/10.5483/bmbrep.2012.45.10.206.

Ballard, D., Winkler-Galicki, J., & Wesoły, J. (2020). Massive parallel sequencing in forensics: Advantages, issues, technicalities, and prospects. *International Journal of Legal Medicine, 134*(4), 1291–1303.

Beecher, D. J. (2006). Forensic application of microbiological culture analysis to identify mail intentionally contaminated with Bacillus anthracis spores. *Applied and Environmental Microbiology, 72*, 5304–5310.

Bocklandt, S., Lin, W., Sehl, M. E., Sanchez, F. J., Sinsheimer, J. S., & Horvath, S. (2011). Epigenetic predictor of age. *PLoS One, 6*, e14821.

Borsteing, C., Fordyce, S. L., Olofsson, J., Mogensen, H. S., & Morling, N. (2014). Evaluation of the ion Torrent™ HID SNP 169-plex: A SNP typing assay developed for human identification by second generation sequencing. *Forensic Science International Genetic, 12*, 144–154.

Børsting, C., & Morling, N. (2015). Next generation sequencing and its applications in forensic genetics. *Forensic Science International Genetic, 18*, 78–89.

Brenig, B., Beck, J., & Schutz, E. (2010). Shotgun metagenomics of biological stains using ultra-deep DNA sequencing. *Forensic Science International Genetic, 4*, 228–231.

Bruijns, B., Tiggelaar, R., & Gardeniers, H. (2018). Massively parallel sequencing techniques for forensics: A review. *Electrophoresis, 39*, 2642–2654.

Chaitanya, L., Breslin, K., Zuñiga, S., Wirken, L., Pośpiech, E., Kukla-Bartoszek, M., Sijen, T., de Knijff, P., Liu, F., Branicki, W., Kayser, M., & Walsh, S. (2018). The HIrisPlex-S system for eye, hair and skin colour prediction from DNA: Introduction and forensic developmental validation. *Forensic Science International Genetic, 35*, 123–135.

Church, G. M. (2006). Genomes for all. *Scientific American, 294*(1), 46–54.

Claes, P., Liberton, D. K., Daniels, K., et al. (2014). Modeling 3D facial shape from DNA. *PLoS Genetics, 10*(3).

Cornelis, S., Gansemans, Y., Deleye, L., Deforce, D., & von Nieuweburhg, R. (2017). Forensic SNP genotyping using nanopore MinION sequencing. *Scientific Reports, 7*(41759).

Courts, C., & Madea, B. (2010). Micro-RNA – a potential for forensic science? *Forensic Science International, 203*, 106–111.

Daniel, R., Santos, C., Phillips, C., Fondevila, M., van Oorschot, R. A., Carracedo, A., Lareu, M. V., & McNevin, D. (2015). A SNaPshot of next generation sequencing for forensic SNP analysis. *Forensic Science International Genetic, 14*, 50–60.

Devesse, L., Ballard, D., Davenport, L., Riethorst, I., Mason-Buck, G., & Syndercombe Court, D. (2017). Concordance of the ForenSeq system and characterisation of sequence-specific autosomal STR alleles across two major population groups. *Forensic Science International Genetic, 34*, 57–61.

Drmanac, R., Drmanac, S., Chui, G., Diaz, R., Hou, A., Jin, H., Jin, P., Kwon, S., Lacy, S., Moeur, B., Shafto, J., Swanson, D., Ukrainczyk, T., Xu, C., & Little, D. (2002). Sequencing by hybridization (SBH): Advantages, achievements, and opportunities. *Advances in Biochemical Engineering/Biotechnology, 77*, 75–101.

Dunisławska, A., Łachmańska, J., Sławińska, A., & Siwek, M. (2017). Next generation sequencing in animal science - a review. *Animal Science Papers And Reports, 35*(3), 205–224.

Eduardoff, M., Santos, C., de la Puente, M., Gross, T. E., Fondevila, M., Strobl, C., Sobrino, B., Ballard, D., Schneider, P. M., Carracedo, A., Lareu, M. V., Parson, W., & Phillips, C. (2015). Interlaboratory evaluation of SNP based forensic identification by massively parallel sequencing using the Ion PGM (TM). *Forensic Science International-Genetics, 17*, 110121.

Eduardoff, M., Xavier, C., Strobl, C., Casas-Vargas, A., & Parson, W. (2017). Optimized mtDNA control region primer extension capture analysis for forensically relevant samples and highly compromised mtDNA of different age and origin. *Genes, 8*(10), 237.

Elaine, Y. Y. C., Elizabeth Gahan, M., & McNevin, D. (2018). Prediction of biogeographical ancestry in admixed individuals. *Forensic Science International: Genetics, 36*, 104–111.

Fierer, N., Lauber, C. L., Zhou, N., McDonald, D., Costello, E. K., & Knight, R. (2010). Forensic identification using skin bacterial communities. *Proceedings of the National Academy of Sciences of the United States of America, 107*, 6477–6481.

Forat, S., Huettel, B., & Reinhardt, R. (2016). Methylation markers for the identification of body fluids and tissues from forensic trace evidence. *PLoS One, 11*(2), e0147973.

Freire-Aradas, A., Phillips, C., & Lareu, M. V. (2017). Forensic individual age estimation with DNA: From initial approaches to methylation tests. *Forensic Science Review, 29*, 121–144.

Freire-Aradas, A., Pospiech, E., ´ Aliferi, A., Girón-Santamaría, L., Mosquera-Miguel, A., Pisarek, A., Ambroa-Conde, A., Phillips, C., Casares de Cal, M. A., Gómez-Tato, A., Spólnicka, M., Wozniak, A., Álvarez-Dios, J., Ballard, D., Court, D. S., Branicki, W., Carracedo, Á., & Lareu, M. V. (2020). A comparison of forensic age prediction models using data from four DNA methylation technologies. *Frontiers in Genetics, 11*, 932.

Frumkin, D., Wasserstrom, A., Budowle, B., & Davidson, A. (2011). DNA methylation-based forensic tissue identification. *Forensic Science International Genetic, 5*, 517−524, 2011.

Fullwood, M. J., Wei, C. L., Liu, E. T., & Ruan, Y. (2009). Next-generation DNA sequencing of paired-end tags (PET) for transcriptome and genome analyses. *Genome Research, 19*, 521−532.

Gettings, K. B., Aponte, R. A., Vallone, P. M., & Butler, J. M. (2015). STR allele sequence variation: Current knowledge and future issues. *Forensic Science International Genetic, 18*, 118−130.

Giampaoli, S., Berti, A., Di Maggio, R. M., Pilli, E., Valentini, A., & Valeriani, F. (2014). The environmental biological signature: NGS profiling for forensic comparison of soils. *Forensic Science International, 240*, 41−47.

Goddard, M. E., & Hayes, B. J. (2009). Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nature Reviews Genetics, 10*, 381−391.

Goodwin, S., McPherson, J. D., & McCombie, W. R. (2016). Coming of age: Ten years of next-generation sequencing technologies. *Nature Reviews Genetics, 17*(06), 333−351.

Grandell, I., Samara, R., & Tillmar, A. O. (2016). A SNP panel for identity and kinship testing using massive parallel sequencing. *International Journal of Legal Medicine, 130*(4), 90514.

Guo, F., Zhou, Y., Liu, F., Yu, J., Song, H., Shen, H., Zhao, B., Jia, F., Hou, G., & Jiang, X. (2016). Next generation sequencing of SNPs using the HID-ion AmpliSeq™ identity panel on the ion torrent PGM™ platform. *Forensic Science International: Genetics, 28*, 82−89.

Hanna, G. J., Johnson, V. A., Kuritzkes, D. R., Richman, D. D., Martinez-Picado, J., Sutton, L., Hazelwood, J. D., & Aquila, R. T. D. (2000). Comparison of sequencing by hybridization and cycle sequencing for genotyping of human immunodeficiency virus type 1 reverse transcriptase. *Journal of Clinical Microbiology, 38*(7), 2715−2721.

Hanson, E., Ingold, S., Haas, C., & Ballantyne, J. (May 2018). Messenger RNA biomarker signatures for forensic body fluid identification revealed by targeted RNA sequencing. *Forensic Science International: Genetics, 34*, 206−221.

Haque, F., Li, J., Wu, H. C., Liang, X. J., & Guo, P. (2013). Solid-state and biological nanopore for real-time sensing of single chemical and sequencing of DNA. *Nano Today, 8*, 56−74.

Harris, B., 1 Hicks, J., Prarat, M., Sanchez, S., & Crossley, B. (2021). Next-generation sequencing capacity and capabilities within the national animal Health laboratory Network. *Journal of Veterinary Diagnostic Investigation, 33*(2), 248−252.

Hayes, B. J., & Daetwyler, H. D. (2019). Bull genomes project to map simple and complex genetic traits in cattle: Applications and outcomes. *Annual Review of Animal Biosciences,* 7, 89−102, 1000.

Henry, R. J. (2012). Next-generation sequencing for understanding and accelerating crop domestication. *Briefings in Functional Genomics, 11*(1), 51−56.

Horvath, S. (2013). DNA methylation age of human tissues and cell types. *Genome Biology, 14*, R115.

Horvath, S., & Raj, K. (2018). DNA methylation-based biomarkers and the epigenetic clock theory of ageing. *Nature Reviews Genetics, 19*, 371−384.

Ingold, S., Dørum, G., Hanson, E., Berti, A., Branicki, W., Brito, P., et al. (2018). Body fluid identification using a targeted mRNA massively parallel sequencing approach — results of a EUROFORGEN/EDNAP collaborative exercise. *Forensic Science International: Genetics, 34*, 105−115.

Jake, M., Robinson, J. M., Pasternak, Z., Mason, C. E., & Elhaik, E. (2021). Forensic applications of microbiomics: A review. *Frontiers in Microbiology, 13*.

Johnson, A. A., Akman, K., Calimport, S. R., Wuttke, D., Stolzing, A., & de Magalhães, J. P. (2012). The role of DNA methylation in aging, rejuvenation, and age-related disease. *Rejuvenation Research, 15*(5), 483−494.

Kaul, S., Koo, H. L., Jenkins, J., et al. (2000). Analysis of the genome sequence of the flowering plantArabidopsis thaliana. *Nature, 408*(6814), 796−815.

Kidd, K. K., et al. (2014a). Progress toward an efficient panel of SNPs for ancestry inference. *Forensic Science International: Genetics, 10*, 23–32.

Kidd, K. K., Pakstis, A. J., Speed, W. C., Lagace, R., Chang, J., Wootton, S., Haigh, E., & Kidd, J. R. (2014b). Current sequencing technology makes microhaplotypes a powerful new type of genetic marker for forensics. *Forensic Science International: Genetics, 12*, 215–224.

King, J. L., LaRue, B. L., Novroski, N. M., Stoljarova, M., Seo, S. B., Zeng, X., Warshauer, D. H., Davis, C. P., Parson, W., Sajantila, A., & Budowle, B. (2014). High-quality and high throughput massively parallel sequencing of the human mitochondrial genome using the Illumina MiSeq. *Forensic Science International: Genetics, 12*, 128–135.

Klimentidis, Y. C., & Shriver, M. D. (2009). Estimating genetic ancestry proportions from faces. *PLoS One, 4*(2), e4460.

Kumar, K. R., Cowley, M. J., & Davis, R. L. (2019). Next-generation sequencing and emerging technologies. *Seminars in Thrombosis and Hemostasis, 45*(7), 661–673.

Lee, H. Y., An, J. H., Jung, S. E., Oh, Y. N., Lee, E. Y., Choi, A., Yang, W. I., & Shin, K. J. (2015). Genome-wide methylation profiling and a multiplex construction for the identification of body fluids using epigenetic markers. *Forensic Science International: Genetics, 17*, 17–24.

Lilje, L., Lillsaar, T., Rä tsep, R., Simm, J., & Aaspo ~ llu, A. (2013). Soil sample metagenome NGS data management for forensic investigation. *Forensic Science International: Genetics, 4*, e35–e36.

Liu, L., Li, Y., Li, S., Hu, N., He, Y., Lin, D., Lu, L., & Law, M. (2012). Comparison of next-generation sequencing systems. *BioMed Research International, 2012*. Article ID 251364.

Liu, M., Li, Y., Yang, Y. F., Yan, Y. W., Liu, F., Li, C. X., Zeng, F. M., & Zhao, W. T. (2020). Human facial shape related SNP analysis in Han Chinese populations. *Yi Chuan, 42*(7), 680–690.

Li, C., Zhao, S., Zhang, N., Zhang, S., & Hou, Y. (2013). Differences of DNA methylation profiles between monozygotic twins' blood samples. *Molecular Biology Reports, 40*, 5275–5280.

Lyons, E. A., Scheible, M. K., Sturk-Andreaggi, K., Irwin, J. A., & Just, R. S. (2013). A high throughput Sanger strategy for human mitochondrial genome sequencing. *BMC Genomics, 14*, 881.

MacCallum, W. G., & Hastings, T. W. (1899). A case of acute endocarditis caused by Micrococcus zymogenes (nov. spec.), with a description of the microorganism. *Journal of Experimental Medicine, 4*, 521–534. https://doi.org/10.1084/jem.4.5-6.521

Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A., Berka, J., Braverman, M. S., Chen, Y. J., Chen, Z., Dewell, S. B., Du, L., Fierro, J. M., Gomes, X. V., Godwin, B. C., He, W., Helgesen, S., Ho, C. H., Irzyk, G. P., & …Rothberg, J. M. (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature, 437*(7057), 376–380. https://doi.org/10.1038/nature03959

McCarthy, J. J., McLeod, H. L., & Ginsburg, G. S. (2013). Genomic medicine: A decade of successes, challenges, and opportunities. *Science Translational Medicine, 5*, 189sr4.

McElhoe, J. A., Holland, M. M., Makova, K. D., Su, M. S., Paul, I. M., Baker, C. H., Faith, S. A., & Young, B. (2014). Development and assessment of an optimized next-generation DNA sequencing approach for the mtGenome using the Illumina MiSeq. *Forensic Science International: Genetics, 13*, 20–29.

McEwen, S. A., Wilson, T. M., Ashford, D. A., Heegaard, E. D., & Kournikakis, B. (2006). Microbial forensics for natural and intentional incidents of infectious disease involving animals. *Rev Sci Tech, 25*, 329–339.

Mikheyev, A. S., & Tin, M. M. (2014). A first look at the Oxford Nanopore MinION sequencer. *Molecular Ecology Resources, 14*, 1097–1102.

Nievergelt, C. M., Maihofer, A. X., Shekhtman, T., Libiger, O., Wang, X., Kidd, K. K., & Kidd, J. R. (2013). Inference of human continental origin and admixture proportions using a highly discriminative, ancestry informative 41-SNP panel. *Investigative Genetics, 4*, 13.

Oliveira, M., & Amorim, A. (2018). Microbial forensics: New breakthroughs and future prospects. *Applied Microbiology and Biotechnology, 102*, 10377–10391. https://doi.org/10.1007/s00253-018-9414-6

Ozsolak, F. (2012). Third-generation sequencing techniques and applications to drug discovery. *Expert Opinion on Drug Discovery, 7*, 231–243.

Pakstis, A. J., Fang, R., Furtado, M. R., Kidd, J. R., & Kidd, K. K. (2012). Minihaplotypes as lineage informative SNPs and ancestry inference SNPs. *European Journal of Human Genetics : European Journal of Human Genetics, 20*(11), 1148–1154.

Paliwal, A., Vaissiere, T., & Herceg, Z. (2010). Quantitative detection of DNA methylation states in minute amounts of DNA from body fluids. *Methods, 52*, 242–247.

Pareek, C. S., Smoczynski, R., & Tretyn, A. (2011). Sequencing technologies and genomesequencing. *Journal of Applied Genetics, 52*, 413–435.

Park, J. L., Kwon, O. H., Kim, J. H., Yoo, H. S., Lee, H. C., Woo, K. M., Kim, S. Y., Lee, S. H., & Kim, Y. S. (2014). Identification of body fluid-specific DNA methylation markers for use in forensic science. *Forensic Science International: Genetics, 13*, 147–153.

Parson, W., Ballard, D., Budowle, B., Butler, J. M., Gettings, K. B., Gill, P., Gusmão, L., Hares, D. R., Irwin, J. A., King, J. L., de Knijff, P., Morling, N., Prinz, M., Schneider, P. M., Van Neste, C., Willuweit, S., & Phillips, C. (2016). Massively parallel sequencing of forensic STRs: Considerations of the DNA commission of the international society for forensic genetics (ISFG) on minimal nomenclature requirements forensic science. *International: Genetics, 22*, 54–63.

Parson, W., & Bandelt, H. J. (2007). Extended guidelines for mtDNA typing of population data in forensic science. *Forensic Science International: Genetics, 1*, 13–19.

Parson, W., Brandstatter, A., Alonso, A., Brandt, N., Brinkmann, B., Carracedo, A., Corach, D., Froment, O., Furac, I., Grzybowski, T., Hedberg, K., Keyser-Tracqui, C., Kupiec, T., Lutz-Bonengel, S., Mevag, B., Ploski, R., Schmitter, H., Schneider, P., Syndercombe-Court, D., … Scheithauer, R. (2004). The EDNAP mitochondrial DNA population database (EMPOP) collaborative exercises: Organisation, results and perspectives. *Forensic Science International, 139*, 215–226.

Parson, W., Huber, G., Moreno, L., Madel, M. B., Brandhagen, M. D., Nagl, S., Xavier, C., Eduardoff, M., Callaghan, T. C., & Irwin, J. A. (2015). Massively parallel sequencing of complete mitochondrial genomes from hair shaft samples. *Forensic Science International: Genetics, 15*, 8–15.

Peck, M. A., Brandhagen, M. D., Marshall, C., Diegoli, T. M., Irwin, J. A., & Sturk-Andreaggi, K. (2016). Concordance and reproducibility of a next generation mtGenome sequencing method for high-quality samples using the Illumina MiSeq. *Forensic Science International: Genetics, 24*, 103–111.

Phan, K., Barash, M., Spindler, X., Gunn, P., & Roux, C. (2020). Retrieving forensic information about the donor through bacterial profiling. *International Journal of Legal Medicine, 134*, 21–29. https://doi.org/10.1007/s00414-019-02069-2

Phillips, C., Fang, R., Ballard, D., Fondevila, M., Harrison, C., Hyland, F., MusgraveBrown, E., Proff, C., Ramos-Luis, E., Sobrino, B., Carracedo, A., FurtadoMR, Syndercombe Court, D., Schneider, P. M., & SNPs, Consortium (2007). Evaluation of the Genplex SNP typing system and a 49plex forensic marker panel. *Forensic Science International-Genetics, 1*(2), 180–185.

Phillips, C., Parson, W., Lundsberg, B., Santos, C., Freire-Aradas, A., Torres, M., Eduardoff, M., Børsting, C., Johansen, P., Fondevila, M., Morling, N., Schneider, P., EUROFORGEN-NoE Consortium, Carracedo, A., & Lareu, M. V. (2014). Building a forensic ancestry panel from the ground up: The EUROFORGEN Global AIM-SNP set. *Forensic Science International: Genetics, 11*, 13–25, 2014 Jul.

Phillips, C., Prieto, L., Fondevila, M., Salas, A., Gomez-Tato, A., & Alvarez-Dios, J. (2009). Ancestry analysis in the 11-M Madrid bomb attack investigation. *PLoS One, 4*, e6583.

Phillips, C., Salas, A., Sánchez, J. J., Fondevila, M., Gómez-Tato, A., Alvarez-Dios, J., Calaza, M., de Cal, M. C., Ballard, D., Lareu, M. V., Carracedo, A., & SNPforID Consortium. (December 2007). Inferring ancestral origin using a single multiplex assay of ancestry-informative marker SNPs. *Forensic Science International: Genetics, 1*(3–4), 273–280.

Poinar, H. N., Schwarz, C., Qi, J., Shapiro, B., Macphee, R. D., & Buigues, B. (2006). Metagenomics to paleogenomics: Large-scale sequencing of mammoth DNA. *Science, 311*, 392–394.

Pop, M., & Salzberg, S. L. (2008). Bioinformatics challenges of new sequencing technology. *Trends in Genetics, 24*, 42–149.

Qin, Y., Schneider, T. M., & Brenner, M. P. (2012). Sequencing by hybridization of long targets. *PLoS One, 7*(5), e35819.

Ronaghi, M., Karamohamed, S., Pettersson, M., Uhlén, M., & Nyrén, P. (1996). Real-time DNA sequencing using detection of pyrophosphate release. *Analytical Biochemistry, 242*(1), 84–89.

Sanchez, J. J., Phillips, C., Børsting, C., Balogh, K., Bogus, M., Fondevila, M., Harrison, C. D., Musgrave-Brown, E., Salas, A., Syndercombe-Court, D., Schneider, P. M., Carracedo, A., & Morling, N. (2006). A multiplex assay with 52 single nucleotide polymorphisms for human identification. *Electrophoresis, 27*, 1713–1724.

Sanger, F., & Coulson, A. R. (1975). A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *Journal of Molecular Biology, 94*(3), 441–448.

Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America, 74*(12), 5463–5467.

Schadt, E. E., Turner, S., & Kasarskis, A. (2010). A window into third-generation sequencing. *Human Molecular Genetics, 19*.

Schmedes, S. E., Woerner, A. E., & Budowle, B. (2017). Forensic human identification using skin microbiomes. *Applied and Environmental Microbiology, 83*, 016722–e1717.

Schneider Peter, M., Prainsack, B., & Kayser, M. (2019). The use of forensic DNA phenotyping in predicting appearance and biogeographic ancestry Deutsches Ärzteblatt international. *Dtsch Arztebl Int, 116*, 873–880.

Schneider, G. F., & Dekker, C. (2012). DNA sequencing with nanopores. *Nature Biotechnology*, (4), 326–328.

Sender, R., Fuchs, S., & Milo, R. (2016). Revised estimates for the number of human and bacteria cells in the body. *PLoS Biology, 14*, e1002533.

Seo, S. B., King, J. L., Warshauer, D. H., Davis, C. P., Ge, J., & Budowle, B. (2013). Single nucleotide polymorphism typing with massively parallel sequencing for human identification. *International Journal of Legal Medicine, 127*(6), 1079–1086.

Simard, A. M., DesGroseillers, L., & Sarafian, V. (2012). Assessment of RNA stability for age determination of body fluid stains. *Canadian Society of Forensic Science Journal, 45*(4), 179–194.

Singh, V. K., Singh, A. K., Singh, S., & Singh, B. D. (2015). Next-generation sequencing (NGS) tools and impact in plant breeding. In J. Al-Khayri, S. Jain, & D. Johnson (Eds.), *Advances in plant breeding strategies: Breeding, Biotechnology and molecular tools*. Cham: Springer.

Slatko, B. E., Gardner, A. F., & Ausubel, F. M. (2018). Overview of next generation sequencing technologies. *Current Protocols in Molecular Biology, 122*(1), e59.

Sobrino, B., & Carracedo, A. (2005). SNP typing in forensic genetics: A review. *Methods in Molecular Biology, 297*, 107–126.

Stoloff, D. H., & Wanunu, M. (2013). Recent trends in nanopores for biotechnology. *Current Opinion in Biotechnology, 24*, 699–704.

Templeton, J. E., Brotherton, P. M., Llamas, B., Soubrier, J., Haak, W., Cooper, A., & Austin, J. J. (2013). DNA capture and next-generation sequencing can recover whole mitochondrial genomes from highly degraded samples for human identification. *Investigative Genetics, 4*(1), 26.

Thompson, J. F., & Milos, P. M. (2011). The properties and applications of single-molecule DNA sequencing. *Genome Biology, 12*, 217.

Titia Sijen, T., & Harbison, S. A. (2021). On the identification of body fluids and tissues: A crucial link in the investigation and solution of crime. *Genes, 12*, 1728.

Van Dijk, E. L., Auger, H., Jaszczyszyn, Y., & Thermes, C. (2014). Ten years of next-generation sequencing technology. *Trends in Genetics, 30*, 418–426.

Vázquez-Baeza, Y., Callewaert, C., Debelius, J., Hyde, E., Marotz, C., & Morton, J. T. (2018). Impacts of the human gut microbiome on therapeutics. *Annual Review of Pharmacology and Toxicology, 58*, 253–270.

Walsh, S., Liu, F., Ballantyne, K. N., van Oven, M., Lao, O., & Kayser, M. (2011). IrisPlex: A sensitive DNA tool for accurate prediction of blue and brown eye colour in the absence of ancestry information. *Forensic Science International: Genetics, 5*, 170–180.

Walsh, S., Liu, F., Wollstein, A., Kovatsi, L., Ralf, A., Kosiniak Kamysz, A., Branickid, W., & Kayser, M. (2013). The HIrisPlex system for simultaneous prediction of hair and eye colour from DNA Forensic Science. *International: Genetics,* 7(1), 98–115.

Wang, M., Wang, Z., He, G., Liu, J., Wang, S., Qian, X., Lang, M., Li, J., Xie, M., Li, C., & Hou, Y. (2019). Developmental validation of a custom panel including 165 Y-SNPs for Chinese Y-chromosomal haplogroups dissection using the ion S5 XL system. *Forensic Science International: Genetics, 38*, 70–76.

Weber-Lehmann, J., Schilling, E., Gradl, G., Richter, D. C., Wiehler, J., & Rolf, B. (2014). Finding the needle in the haystack: Differentiating "identical" twins in paternity testing and forensics by ultra-deep next generation sequencing. *Forensic Science International: Genetics, 9*, 42–46.

Wendt, F. R., King, J. L., Novroski, N. M. M., Churchill, J. D., Ng, J., Oldt, R. F., McCulloh, K. L., Weise, J. A., Smith, D. G., Kanthaswamy, S., & Budowle, B. (2017). Flanking region variation of ForenSeq DNA signature Prep kit STR and SNP loci in Yavapai native Americans. *Forensic Science International: Genetics, 28*, 146–154.

Wold, B., & Myers, R. M. (2008). Sequence census methods for functional genomics. *Nature Methods, 5*, 19–21.

Wooley, J. C., Godzik, A., & Friedberg, I. (2010). A primer on metagenomics. *PLoS Computational Biology, 6*(2), e1000667.

Xavier C., de la Puente M., Mosquera-Miguel A., Freire-Aradas A., Kalamara V., Vidaki A., Gross T., Revoir A., Pospiech E., Kartasinska E., Spolnicka M., Branicki W., Ames C. E., Schneider P. M., Hohoff C., Kayser M., Phillips C., Parson W., On Behalf of the VISAGE Consortium. (2020). Development and validation of the VISAGE AmpliSeq basic tool to predict appearance and ancestry from DNA. *Forensic Science International: Genetics*, 48, 102336

Xu, H., Zhao, Y., Liu, Z., Zhu, W., Zhou, Y., & Zhao, Z. (2012). Bisulfite genomic sequencing of DNA from dried blood spot microvolume samples. *Forensic Science International: Genetics, 6*, 306–309.

Xue, Y., Wang, Q., Long, Q., Ng, B. L., Swerdlow, H., Burton, J., Skuce, C., Taylor, R., Abdellah, Z., Zhao, Y., Asan, MacArthur, D. G., Quail, M. A., Carter, N. P., Yang, H., & Tyler-Smith, C. (2009). Human Y chromosome base-substitution mutation rate measured by direct sequencing in a deep-rooting pedigree. *Current Biology, 19*, 1453–1457.

Yang, M. Q., Athey, B. D., Arabnia, H. R., Sung, A. H., Liu, Q., Yang, J. Y., Mao, J., & Deng, Y. (2009). High-throughput next-generation sequencing technologies foster new cutting-edge computing techniques in bioinformatics. *BMC Genomics, 10*(Suppl. 1).

Yang, X., Jiao, R., Yang, L., Wu, L. P., Li, Y. R., & Wang, J. (2011). New-generation high-throughput technologies based 'omics' research strategy in human disease. *Yi Chuan, 33*, 829–846.

Yang, Y., Xie, B., & Yan, J. (2014). Application of next-generation sequencing technology in forensic science. *Genomics, Proteomics & Bioinformatics, 12*, 190–197.

Yu, J., Hu, S., Wang, J., et al. (2002). A draft sequence of the rice genome (Oryza sativa L. ssp. indica). *Science, 296*(5565), 79–92.

Zhang, S., Bian, Y., Chen, A., Zheng, H., Gao, Y., Hou, Y., & Li, C. (2017). Developmental validation of a custom panel including 273 SNPs for forensic application using Ion Torrent PGM. *Forensic Science International: Genetics, 27*, 50–57.

Zubakov, D., Hanekamp, E., Kokshoorn, M., van Ijcken, W., & Kayser, M. (2008). Stable RNA markers for identification of blood and saliva stains revealed from whole genome expression analysis of time-wise degraded samples. *International Journal of Legal Medicine, 122*(2), 135–142.

CHAPTER 4

Processing of biological samples for forensic NGS analysis

Sulekha Coticone and Camila Garcia
Department of Chemistry and Physics, Florida Gulf Coast University, Ft Myers, FL, United States

Overview of next-generation sequencing

Genomic research has been revolutionized by next-generation sequencing (NGS), which allows complete genomes to be sequenced in a single day at a reduced cost (Behjati et al., 2013; Van Djik et al., 2014). This has resulted in significant improvements in disease diagnosis, prognosis, and treatment, as well as answers to genetic issues from a variety of applications, including forensic science (Fig. 4.1). NGS is now a must-have tool for any scientist. From quick single nucleotide polymorphism (SNP) genotyping of a single individual to whole genome sequencing (WGS) of huge populations, ultrahigh-throughput NGS technologies have a wide range of applications and are fully scalable. All NGS platforms perform the sequencing of millions of small fragments of DNA in parallel. Bioinformatics analyses are then used to piece together these fragments by mapping the individual reads to the human reference genome.

NGS versus Sanger sequencing: What's the difference?

The concepts behind Sanger and NGS are identical in theory. DNA polymerase adds fluorescent nucleotides one by one to a developing DNA template strand in both NGS and Sanger sequencing (also known as dideoxy or capillary electrophoresis (CE) sequencing). The fluorescent tag on each integrated nucleotide is used to identify it. The volume of sequencing differs significantly between Sanger and NGS. While the Sanger method only sequences a single DNA fragment at a time, NGS runs can sequence millions of fragments at the same time. This procedure results in the simultaneous sequencing of hundreds to thousands of genes. With deep sequencing, NGS also has a higher discovery capability for detecting novel or rare variants.

Next-generation sequencing in forensic science

NGS technology has the potential to be used in the field of forensic science as shown in Fig. 4.1 DNA database creation, ancestry, phenotypic inference, monozygotic twin investigations, bodily fluid, species identification, forensic animal, plant, and microbiological

Next Generation Sequencing (NGS) Technology in DNA Analysis
ISBN 978-0-323-99144-5, https://doi.org/10.1016/B978-0-323-99144-5.00028-7
© 2024 Elsevier Inc.
All rights reserved.

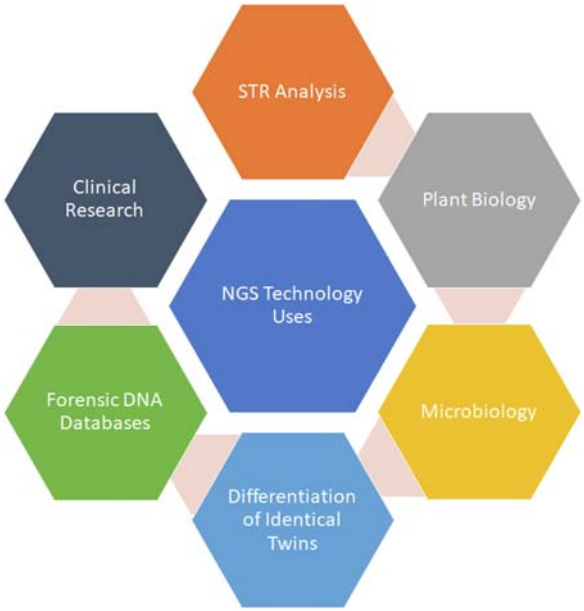


Figure 4.1 NGS technology uses.

analyses are only a few examples (Weber-Lehmann et al., 2014; Yang et al., 2014). NGS not only facilitates the identification of mixed DNA samples and the analysis of complex paternity cases but also has potential advantages for short tandem repeat (STR) testing. These include high throughput, low cost, simultaneous detection of large numbers of STR loci, and the ability to distinguish alleles with similar length or digital read count. Forensic experts are therefore actively looking into ways to transition from CE to NGS for DNA profile analysis. NGS has various advantages over CE, including the ability to combine data from STR and SNP loci, small amplicons without size separation limits, higher discrimination power, deep mixture resolution, and sample multiplexing (Van Neste et al., 2014). In a single, tailored assay, NGS allows examiners to generate data that spans the human genome and answers a larger range of questions. Furthermore, the STR calls generated by NGS are perfectly compatible with current database formats. NGS can also examine forensic mitochondrial DNA and target the SNP marker sets not commonly available with classic CE-based approaches. Studies show that NGS may be used to analyze even the most degraded, and highly mixed evidential samples with encouraging results.

Overview of the sample preparation process for next-generation sequencing

Several steps must be completed in order to analyze samples using NGS. Fig. 4.2 depicts an overview of the steps in the NGS sample preparation process. Sample handling and



Figure 4.2 Overview of the steps in the NGS sample preparation process.

processing are the first steps, followed by DNA extraction and quantitation. After extracting the DNA samples and determining their quality and quantity, the library preparation process begins. Prior to multiplexing and denaturation, samples undergo purification and normalization processes following library preparation.

Sample preparation methods

Genetic material can be extracted from a number of biological samples, including blood, cultured cells, biopsies, tissue sections, teeth, bone, and urine, as well as microbes or plant specimens, depending on the goal of the investigation (Carrasco et al., 2020). The samples used for NGS forensic DNA typing could be ancient or they could be immaculate reference samples. Sample handling is crucial during NGS since samples may be degraded or in limited quantities. Additionally, due to the high sensitivity of the NGS process, care must be taken to prevent contamination. Methods to decrease or eliminate contamination require surfaces and nonautoclavable equipment to be cleaned with the following items/equipment: soap and water, UV light irradiation, bleach (10%), ethanol (70%), or commercially available solutions, e.g., LookOut DNA Erase Spray (Sigma Aldrich), DNAZap (Invitrogen). Many techniques for obtaining DNA from bone and tooth samples have been investigated. Because the DNA is preserved in bone cells, marrow, and tooth pulp, the samples must first be cleaved, decalcified, and drilled. Crushed samples are reduced to tiny bits or powder. To avoid carryover of dust between samples and the work area, crushing and pulverizing bone and teeth are done using a laminar flow hood. Between uses, stainless steel equipment, such as freezer mill components and die presses used to crush and pulverize bone and teeth, is cleaned and autoclaved. To decrease the risk of contamination while extracting DNA from bones, all of the activities were carried out while wearing face masks, and all reagents and consumables were properly sterilized using short-wave ultraviolet light, bleach, or both before leaving the clean room (Gaudio et al., 2019). The petrous bone samples were first cleaned by immersing them in 5% bleach for 30 s, then rinsed with ethanol and dried. Bone powder was then extracted from the cochlea using a direct drilling (Dremel 9100 Fortiflex) rotary tool fitted with a small-sized spherical 1.5 mm grinding bit from the inferior side of the petrous portion, which was contaminated with bleach and ethanol.

DNase- and RNase-free tubes and aerosol filter tips are used while working with chemicals and samples. Samples should be processed in a separate preparation area from the amplification area. When sanding and crushing skeletal remains, analysts should use N95 protective masks to avoid dust inhalation. When processing materials in the DNA lab, other personal protection equipment such as goggles, gloves, hair nets, masks, lab coats, and closed-toed shoes should always be used.

DNA extraction

For a successful NGS, the quality and amount of the extracted samples are also critical. As a result, before moving on to the next phases, attention must be paid to selecting a suitable extraction method and setting quality control parameters. Success on any NGS platform depends on optimal sample preparation. The rising demand for NGS typically puts pressure on upstream processes to analyze more samples and prepare high-quality DNA for library prep and analysis.

There are various ways of extracting DNA from cellular material. While organic phenol-chloroform-isoamyl alcohol (PCI or PCIA) (Butler, 2005) and inorganic metal-chelating Chelex-100 resin (Walsh et al., 1991) were formerly popular for forensic DNA extraction, silica-based approaches currently predominate (Brevnov et al., 2009; Castella et al., 2006; Elkins et al., 2021; Eychner et al., 2017; Hoff-Olsen et al., 1999). The PCIA method remains the gold standard for most applications because it yields the most intact DNA from samples. The caveat is that all phenolic organic compounds must be removed to prevent the DNA phosphodiester bonds from being attacked by the phenolic hydroxyl group, resulting in fragmented, degraded DNA. Furthermore, the PCIA technique is not appropriate for all materials; for example, it digests chewing gum (Eychner et al., 2017). The Chelex-100 technique is a low-cost, hands-on alternative method, but it, like the PCIA method, requires an overnight incubation step that must be accommodated in the lab's standard operating procedure.

The majority of contemporary DNA extraction techniques used in commercial kits are silica-based. Magnetic bead suspensions are used in the Qiagen QIAamp DNA Investigator and EZ1 DNA Investigator kits. Magnetic beads and magnetite-modified silicon oxide magnetic beads are used in other techniques, such as the Applied Biosystems PrepFiler and the Promega DNA IQ kits. Recent research found that the PrepFiler BTA approach outperformed a PCI-silica-based method for extracting DNA from bone (Hasap et al., 2019). These procedures are quick, can be automated, and produce the most repeatable DNA yields. Beckman Coulter Biomek liquid handling platform, Tecan HID EVOlution, Promega Maxwell and Maxprep Instruments, and Qiagen QIAcube HT, QIAcube Connect, and EZ1 instruments are among the robotic equipment available. In a recent investigation, certain crude bone lysates were shown to hinder DNA purification when utilizing paramagnetic silica beads in the DNA IQ Casework Pro Kit and the Maxwell 16 owing to filter clogging, although this was improved when the lysates were treated with phenol (Desmyter et al., 2017). The Qiagen EZ1 robot can be used in conjunction with the EZ1 DNA Investigator Kit. Samples collected on FTA cards or similar preservative-containing fibrous material can be extracted using one of the aforementioned kits or techniques, such as the EZ1 DNA Investigator Kit, or put straight into the polymerase chain reaction (PCR) sample tube in library creation (Kampmann et al., 2016). A modified process using the Qiagen EZ1 DNA Investigator

Kit, silica-based technology, and the EZ1 BioRobot functionality for extracting DNA from current and historic bone for use in DNA typing analysis (Dukes et al., 2012; Klavens et al., 2021). Given that contemporary human DNA typing kits generally require 1 ng of input DNA, any of the extraction procedures should be able to recover enough DNA from blood or buccal samples.

For DNA from chemically treated or degraded bone samples, the Armed Forces Medical Examiner System's Armed Forces DNA Identification Laboratory has developed a protocol using demineralized samples (Marshall et al., 2017). Briefly, DNA was isolated from 0.2 to 1.0 g powdered bone sample. The bone powder was first demineralized in a buffer containing 0.5 M EDTA, 1% sarkosyl, and 20 mg/mL proteinase K by incubating overnight with agitation at 56°C. DNA was purified using one of three protocols: organic, QIAquick PCR Purification Kit (QIAGEN, Hilden, Germany), or a variation of the procedure that utilized the MinElute PCR Purification Kit (QIAGEN) instead of the QIAquick (Loreille et al, 2007, 2010).

When four DNA extraction techniques were assessed (Edson, 2019), it was revealed that the best DNA recovery from postcranial osseous human remains of military soldiers lost in World War II, the Korean War, and Southeast Asia was achieved by employing a comprehensive demineralization approach with organic extraction. Zeng et al. (2019) investigated the DNA IQ, DNA Investigator, and PrepFiler BTA kits, as well as the organic extraction method for hair and blood and two total demineralization protocols for bone; they discovered that all of the DNA extraction methods were efficient and compatible with the Precision ID and ForenSeq kits. Carrasco et al. (2020) described a customized extraction procedure for samples of deteriorated blood and dental remnants. Briefly, genomic DNA was extracted from dental tissue samples using the one-step, 1-h QuickExtract FFPE DNA Extraction Kit (Lucigen) according to the manufacturer's instructions. A blood sample was extracted using organic methods (Proteinase K (ProK) and SDS digestion with phenol-chloroform extraction) and purified by cold ethanol precipitation followed by resuspension in Tris-EDTA buffer. Degraded blood samples were created by subjecting the extracted DNA to sonication at 50°C for time spans ranging from 0 to 16 h. Sidstedt et al. (2020) investigated PCR inhibition in massively parallel sequencing (MPS) applications. The authors mention that though the classical solution to handle PCR inhibition is to purify or dilute DNA extracts, which leads to DNA loss, inhibitor-tolerant DNA polymerases either single enzymes or blends provide a more straightforward solution.

DNA recovery from trace materials such as hair, fingernails, and fingerprints, as well as other contact DNA, has been studied for MPS applications (Naue et al., 2020). Preuner et al. (2014) discovered that using the PrepFiler Forensic DNA Extraction Kit resulted in high-quality DNA extracted from fingernail clippings. Tasker et al. (2017) recovered DNA from improvised explosive device (IED) pipe bombs after they had been detonated using cotton swabs treated with 2% sodium dodecyl sulfate and by the QIAamp DNA Mini Kit (Qiagen Inc., Hilden, Germany). England et al. (2020) used DNA isolated

from laser-microdissected cells to evaluate the ForenSeq DNA Signature Prep Kit. Eychner et al. (2017) evaluated five DNA extraction techniques for recovering DNA from chewing gum and saliva aliquoted on swabs (PCIA, QIAamp DNA Investigator, DNA IQ, Chelex-100, and PrepFiler); the QIAamp method performed the best overall. To concentrate the recovered DNA from low-quantity samples such as chewing gum, touch DNA, and human remains, ethanol precipitation, concentrator devices, and lower elution volumes can be used (Eychner et al., 2017).

For extracting DNA from soil samples, equal amounts of soil were mixed with freshly prepared (to avoid bacterial contamination) saturated phosphate buffer to remove any PCR inhibitors (mainly humic acid). The mixture was then subjected to centrifugation. The supernatant containing the extracellular DNA is then extracted using a commercial kit for soil DNA (NucleoSpin Soil; Macherey-Nagel, Duren, Germany) following the manufacturer's instructions (Giampaoli et al., 2014). In a pilot study for the validation of the Precision ID GlobalFiler NGS STR panel v2 with the Ion S5 system, DNA from samples (blood stain, nail, hair, and buccal swab) was extracted using the DNA extraction kit from Finegene Biotech, Co., Ltd., China (Tao et al., 2019). The DNA from muscle samples was extracted using the DNeasy Blood and Tissue kit (Qiagen).

DNA quantitation

There are numerous DNA quantification methods available for determining the amount of DNA recovered and how much of the extract to use for NGS typing. Human-specific and nonspecific quantification techniques are available. UV-Vis and fluorescence spectroscopy are two spectroscopic approaches for measuring DNA. Spectroscopic techniques for estimating total DNA in a sample can be utilized, although they are not human-specific (Elkins, 2013). Because they need as little as 1 μL of material, the NanoDrop (UV-Vis spectroscopy) and Qubit (fluorescence spectroscopy) spectrophotometers are commonly used for fast DNA measurement. Real-time PCR techniques can be used to determine the amount and amplification of human DNA (Horsman et al., 2006). Commercial assays as well as "home brew" tests can be used. Total human and male genomic DNA is quantified using a real-time PCR technique targeting the TPOX region (Horsman et al., 2006). Several commercial kits are available to measure the amount of human DNA in a sample, as well as the presence of human and male genomic DNA as well as identify PCR inhibition utilizing real-time PCR and multiple dye channel fluorescence detection. These include the Applied Biosystems Quantifiler Human DNA Quantification, Quantifiler DUO, and Quantifiler Trio kits (Barbisin et al., 2009; Green et al., 2005), as well as the Promega Plexor HY and PowerQuant System kits (Krenke et al., 2008). The Qiagen Investigator Quantiplex Pro and HYres Kits measure the total amount of human and male DNA in a sample, identify PCR inhibition, and offer a degradation index (Morrison et al., 2020; Vraneš et al., 2017). While the Quantifiler

and Plexor HY kits take around 2 h to complete, the Quantiplex HYres Kit takes just 1 h. The sensitivity of commercial assays has increased over time, and the most recent kit, the Quantiplex HYres kit, is the most sensitive. The Quantiplex HYres Kit is sensitive to 1 pg/L of total DNA, whereas the Plexor HY kit is sensitive to 6.4 pg. The Quantifiler DUO kit detects 51 pg of DNA. For usage with the PowerSeq 45GY NGS kit, the PowerSeq Quant MS System, QuantiFluor ONE dsDNA System, and QuantiFluor dsDNA System are suggested. The more degraded the DNA, the fewer loci should be predicted to be typed using any DNA typing method. A precise measurement of the amount of human DNA in a sample is required for estimating the proper eluent input for NGS. The first step with NGS is to prepare a library. The extracted DNA can be diluted to obtain the ideal input quantity specified by the manufacturer of the NGS kit. Low template samples may be utilized and input improved by adding more extract and no water to the library preparation PCR reactions, but if sufficient DNA is not given in the reaction, a lower number of reads and coverage may result. Pajnič et al. (2020) used the PowerQuant kit to quantify human remains from a World War II slaughter of a Slovenian family, producing DNA profiles from a molar and five femurs.

In a study where the beta version of the ForenSeq DNA Signature Prep Kit MiSeq FGx system was evaluated, casework-type and ancient DNA samples were extracted and then quantitated with a TaqMan assay targeting an Alu repetition motif of 79 bp (Niederstatter et al., 2007; Xavier et al., 2017). For the developmental validation of the MiSeq FGx forensic genomics system, buccal cell samples were quantitated using the Quantifiler Human DNA quantification kit (Life Technologies, Carlsbad, CA) on the Stratagene Mx3000P qPCR System (Agilent, Santa Clara, CA) using the Qubit dsDNA HS or BR assay kit on a Qubit 2.0 Fluorometer (Life Technologies), according to the respective manufacturers' instructions (Jäger et al., 2017). In another study, the performance of the MPS-STR panel (Precision ID GlobalFiler and NGS-STR panel (Thermo Fisher Scientific) designated for MPS on the Ion Torrent platform was evaluated using human blood samples that were extracted using the PureLink Genomic DNA Mini kit (Thermo Fisher Scientific), followed by quantitation using the Quantifiler Human DNA quantification kit on an Applied Biosystems 7500 Real-Time PCR system following the manufacturer's instructions (Wang et al., 2017).

Library preparation

Library preparation is a set of processes used to prepare a sample for NGS sequencing. It starts with previously extracted, quantitated, and diluted samples from various sources mentioned above. The number of steps to prepare the library varies substantially depending on the manufacturer and procedure. NGS kits are available from Thermo Fisher, Verogen, Promega, and Qiagen. When utilizing kits from Thermo Fisher Scientific, there are three panels for analyzing STR and SNP markers. These include the Precision

ID GlobalFiler NGS STR panel, the Precision ID Ancestry panel, and the Precision ID Identity panel. The Precision ID GlobalFiler NGS STR Panel includes primer sets for 35 markers, including the same 21 autosomal STR markers, the amelogenin sex markers, and 14 additional informative markers for forensic analysis (Muller et al., 2018; Wang et al., 2017). The Precision ID Ancestry Panel from Applied Biosystems consists of 165 autosomal markers (that provide biogeographic ancestry information), including 55 markers created by Dr. Kenneth Kidd (Kidd et al., 2012) and 123 markers based on a publication by Dr. Michael Seldin (Kosoy et al., 2009). The Precision ID Identity Panel consists of 124 autosomal SNPs with high heterozygosity and a low fixation index (Fst) (Pakstis et al., 2010).

The MiSeq FGx Forensic Genomics System, comprised of the ForenSeq DNA Signature Prep Kit, MiSeq FGx Reagent Kit, MiSeq FGx instrument, and ForenSeq Universal Analysis Software, uses PCR to simultaneously amplify up to 231 forensic loci in a single multiplex reaction (Jäger et al., 2017). Targeted loci include amelogenin, 27 common, forensic autosomal STRs, 24 Y-STRs, 7 X-STRs, and 3 classes of SNPs. The ForenSeq kit includes two primer sets: Amelogenin, 58 STRs, and 94 identity-informative SNPs (iiSNPs) are amplified using DNA Primer Set A (DPMA; 153 loci); if a laboratory chooses to generate investigative leads using DNA Primer Set B, amplification is targeted to the 153 loci in DPMA plus 22 phenotypic-informative SNPs (piSNPs) and 56 biogeographical ancestry-informative SNPs (aiSNPs).

The ForenSeq Signature Prep kit from Verogen uses PCR to simultaneously amplify 231 forensic loci in a single multiplex reaction (Jäger et al., 2017). Targeted loci include amelogenin, 27 common, forensic autosomal STRs, 24 Y-STRs, 7 X-STRs, and 3 classes of SNPs. The ForenSeq kit includes two primer sets: Amelogenin, 58 STRs, and 94 iiSNPs are amplified using DNA Primer Set A (DPMA; 153 loci). If investigative leads are needed using DNA Primer Set B, amplification is targeted to the 153 loci in DPMA plus 22 piSNPs and 56 biogeographical aiSNPs. The ForenSeq MainstAY kit from Verogen contains 53 standard loci for use in forensic analysis, relationship testing, and research. The kit combines 27 autosomal STRs, 25 YSTRs, and amelogenin in a single reaction to provide higher likelihood ratios and enable direct comparison of single-source samples.

The Promega PowerSeq 46GY kit is a 46-locus multiplex system that targets 22 autosomal STR markers, 23 Y-STRs, and amelogenin. The extended panel of STR markers is intended to satisfy both the Combined DNA Index System and the European Standard Set recommendations.

While many NGS applications require WGS, forensic genomic applications use a tailored sequencing technique that uses an amplicon-based workflow to target forensically relevant sequences. The library preparation begins with amplifying and enriching the targets, followed by the addition of tags (tagmentation), indices for demultiplexing, and adapter sequences for flow cell binding, and finally library purification and

normalization. During the library preparation process, in addition to the samples to be sequenced, a positive control (e.g., 2800 M) and a negative control (no template control) are processed. The ForenSeq kit requires a two-step amplification to prepare the library. For each forensically important target sequence in the DNA sample, the first PCR step uses sequence-specific, tagged primer pairs. Indexes and adapters are integrated into the amplicons during the second PCR cycle. After that, the amplicon libraries are purified, pooled, and linearized in a single tube. During the first PCR step, a forward tag connected to the forward primer and a reverse tag linked to the reverse primer are added to the target amplicon. All of the targets have the same sequence tags appended to them. Using the tags added in the first step of PCR, adapter sequences are added adjacent to the primer sequences to allow the amplicons to bind to the flow cell for sequencing. The unique forward and reverse adapter i7 and i5 index sequence combinations are added to the targets to label them for demultiplexing interpretation in a second PCR step. The library is injected into a flow cell, where fragments are captured on a lawn of surface-bound oligos complementary to the library adapters. The indices are a distinct collection of sequences used to allocate data to samples. An index, which is similar to a barcode, is made up of an 8-bp sequence that is used to demultiplex the sample data once it has been sequenced. As a result, the completed "library" for each target includes an i5 adapter, an i5 index, a forward tag, a target sequence, a reverse tag, an i7 index, and an i7 adapter. The ForenSeq first and second PCR steps produce amplicons ranging in size from 60 to 460 bp. The French National Police used a Hamilton ID NGS-V STARlet robot to automate the library preparation procedure (Laurent et al., 2017).

The Thermo Fisher Ion Chef robot can be used to perform NGS library preparation for any of the human identification (HID) Ion AmpliSeq panels, including the GlobalFiler NGS STR Panel v2, Identity Panel, AmpliSeq Identity Panel, and the Ampliseq Ancestry Panel. When employing the robot, the plate containing the eight samples to be sequenced, consumables, and the master mix are all loaded into the Ion Chef. Consumables include Ion S5 Precision ID Chef Solutions reagent cartridges, a chip adapter, an enrichment strip cartridge, a tip cartridge with pipet tips, a PCR plate and frame seal, a disposable lid for the recovery station, recovery tubes, and one or two sequencing chips. The barcodes of the objects placed in the instrument are scan by a camera system. The Ion Chef contains a thermocycler that automates the library preparation, library purification, library normalization, and chip loading procedures for the eight samples in a 7-h process. Using a combination of restriction enzymes, each DNA sample is sliced into millions of pieces. The Ion Xpress Barcode Adapters 1−96 Kit and the IonCode Barcode Adapters 1−384 Kit are used in the kit. Barcoded libraries may be concatenated and loaded onto a single Ion chip, reducing sequencing run time and cost while still allowing for reliable sample-to-sample comparisons. Each fragment binds to its own primer-coated bead, giving rise to template ion sphere particles (ISPs). The ISPs are cleansed, and those that test positive for the template are put into the chip at the enrichment

station. Each bead falls into a well after flowing across the chip. When employing the Ion Chef with a 530 chip, 24 samples may be put on one chip in around 10 h for autosomal NGS library creation.

To summarize, a library is a DNA sample that has indexes and adapters connected to either end and is ready for sequencing. Following library preparation, the samples are mixed together in the same tube to create a mixture, or multiplex, of different libraries pooled together and ready for sequencing.

Library normalization

The library purification process comes after library preparation. Excess primers and reagents are removed during the library purification or clean-up process. The ForenSeq Signature Prep Kit makes use of magnetic beads. Working with fewer samples at a time during the bead-based phases results in better sequencing findings. The library is initially added to a consistent number of magnetic beads in the method, and the DNA binds the beads. A magnetic block at the bottom of the plate is utilized to attract the DNA-bound beads to its surface. Ethanol washes are used to clean the beads and eliminate any leftover primers for PCR reagents. After removing the extra ethanol, the DNA library is freed from the beads using a resuspension buffer. Users must use caution to ensure that all of the ethanol is removed. The library produced following the purification stage contains more of each library than will be required for sequencing. Prior to sequencing, the amplicon lengths and quality can be checked using an agarose or polyacrylamide gel, a BioAnalyzer, or XIAxcel equipment. High-quality ForenSeq libraries have amplicons in the 60–460 bp range. Primer dimers are identified by short amplicons of about 60 bp.

As previously stated, the Ion Chef handles not only library preparation but also library quantification, purification, library normalization, and chip loading. The library normalization phase comes after library purification. The Ion Library TaqMan Quantitation Kit may be used to quantify the eight-sample library. Because the PCR processes in library preparation can provide a wide range of yields, the library normalization step is used to equalize the quantity and concentration of each sample library to guarantee that each library is represented equally when pooled. Following the determination of which samples will be sequenced in the same run, the generated libraries must be normalized and pooled together for the sequencing run. Libraries generated at separate periods can be normalized together to be sequenced on the same flow cell if they contain unique index combinations. The pooled libraries are diluted to 50 pM and combined in the order of the 530 chip's barcode adapters (1–32).

Bead-based normalization is used in the ForenSeq Signature Prep Kit. The normalizing procedure employs a consistent number of magnetic beads to bind the same amount

of library to each sample. Before usage, the beads must be warmed to room temperature and well mixed to guarantee an equitable distribution of beads to each library. Beads are introduced to each well containing a library and bind an equal, maximum amount of each library depending on the binding capacity and number of beads added. The magnetic beads bind the DNA libraries, remove the surplus, and then elute the normalized libraries from the beads.

According to Verogen, the library preparation stage utilizing the ForenSeq Signature Prep Kit takes around 9 h. In practice, the hands-on time varies depending on the number of samples and the investigator's competence in creating the library with the processes and methods.

References

Barbisin, M., Fang, R., O'Shea, C. E., Calandro, L. M., Furtado, M. R., & Shewale, J. G. (2009). Developmental validation of the quantifiler® duo DNA quantification kit for simultaneous quantification of total human and human male DNA and detection of PCR inhibitors in biological samples*. *Journal of Forensic Sciences, 54*(2), 305–319. https://doi.org/10.1111/j.1556-4029.2008.00951.x

Behjati, S., & Tarpey, P. S. (2013). What is next generation sequencing? Archives of disease in childhood. *Education and Practice Edition, 98*(6), 236–238. https://doi.org/10.1136/archdischild-2013-304340

Brevnov, M. G., Pawar, H. S., Mundt, J., Calandro, L. M., Furtado, M. R., & Shewale, J. G. (2009). Developmental validation of the PrepFiler™ forensic DNA extraction kit for extraction of genomic DNA from biological samples*. *Journal of Forensic Sciences, 54*(3), 599–607. https://doi.org/10.1111/j.1556-4029.2009.01013.x

Butler, J. M. (2005). *Forensic DNA typing: Biology, technology, and genetics of STR markers*. Elsevier.

Carrasco, P., Inostroza, C., Didier, M., Godoy, M., Holt, C. L., Tabak, J., & Loftus, A. (2020). Optimizing DNA recovery and forensic typing of degraded blood and dental remains using a specialized extraction method, comprehensive QPCR sample characterization, and massively parallel sequencing. *International Journal of Legal Medicine, 134*(1), 79–91. https://doi.org/10.1007/s00414-019-02124-y

Castella, V., Dimo-Simonin, N., Brandt-Casadevall, C., & Mangin, P. (2006). Forensic evaluation of the QIAshredder/QIAamp DNA extraction procedure. *Forensic Science International, 156*(1), 70–73. https://doi.org/10.1016/j.forsciint.2005.11.012

Desmyter, S., De Cock, G., Moulin, S., & Noël, F. (2017). Organic extraction of bone lysates improves DNA purification with silica beads. *Forensic Science International, 273*, 96–101. https://doi.org/10.1016/j.forsciint.2017.02.003

Dukes, M. J., Williams, A. L., Massey, C. M., & Wojtkiewicz, P. W. (2012). Technical note: Bone DNA extraction and purification using silica-coated paramagnetic beads. *American Journal of Physical Anthropology, 148*(3), 473–482. https://doi.org/10.1002/ajpa.22057

Edson, S. M. (2019). Extraction of DNA from skeletonized postcranial remains: A discussion of protocols and testing modalities. *Journal of Forensic Sciences, 64*(5), 1312–1323. https://doi.org/10.1111/1556-4029.14050

Elkins, K. M. (2013). *Forensic DNA biology: A laboratory manual*. Academic Press.

Elkins, K. M., & Zeller, C. B. (2021). *Next generation sequencing in forensic science: A primer*. CRC Press. https://doi.org/10.4324/9781003196464

England, R., Nancollis, G., Stacey, J., Sarman, A., Min, J., & Harbison, S. (2020). Compatibility of the ForenSeq™ DNA signature prep kit with laser microdissected cells: An exploration of issues that arise with samples containing low cell numbers. *Forensic Science International: Genetics, 47*, 102278. https://doi.org/10.1016/j.fsigen.2020.102278

Eychner, A. M., Schott, K. M., & Elkins, K. M. (2017). Assessing DNA recovery from chewing gum. *Medicine, Science & the Law, 57*(1), 7–11. https://doi.org/10.1177/0025802416676413

Gaudio, D., Fernandes, D. M., Schmidt, R., Cheronet, O., Mazzarelli, D., Mattia, M., O'keefe, T., Feeney, R. N. M., Cattaneo, C., & Pinhasi, R. (2019). Genome-wide DnA from degraded petrous bones and the assessment of sex and probably geographic origins of forensic cases. *Nature Scientific Reports, 9*(8226), 1–11. https://doi.org/10.1038/s41598-029-44638-w

Giampaoli, S., Berti, A., DiMaggio, R. M., Pili, E., Valentini, A., Valeriani, F., Gianfranceschi, G., Barni, F., Ripani, L., & Romano Spica, V. (2014). The environmental biological signature: NGS profiling for forensic comparison of soils. *Forensic Science International, 240*, 41–47.

Green, R. L., Roinestad, I. C., Boland, C., & Hennessy, L. K. (2005). Developmental validation of the quantifiler™ real-time PCR kits for the quantification of human nuclear DNA samples. *Journal of Forest Science, 50*(4), 809–825.

Hasap, L., Chotigeat, W., Pradutkanchana, J., Asawutmangkul, W., Kitpipit, T., & Thanakiatkrai, P. (2019). Comparison of two DNA extraction methods: PrepFiler® BTA and modified PCI-silica based for DNA analysis from bone. *Forensic Science International: Genetics Supplement Series,* 7(1), 669–670. https://doi.org/10.1016/j.fsigss.2019.10.132

Hoff-Olsen, P., Mevåg, B., Staalstrøm, E., Hovde, B., Egeland, T., & Olaisen, B. (1999). Extraction of DNA from decomposed human tissue: An evaluation of five extraction methods for short tandem repeat typing. *Forensic Science International, 105*(3), 171–183. https://doi.org/10.1016/S0379-0738(99)00128-0

Horsman, K. M., Hickey, J. A., Cotton, R. W., Landers, J. P., & Maddox, L. O. (2006). Development of a human-specific real-time PCR assay for the simultaneous quantitation of total genomic and male DNA*. *Journal of Forensic Sciences, 51*(4), 758–765. https://doi.org/10.1111/j.1556-4029.2006.00183.x

Jäger, A. C., Alvarez, M. L., Davis, C. P., Guzmán, E., Han, Y., Way, L., Walichiewicz, P., Silva, D., Pham, N., Caves, G., Bruand, J., Schlesinger, F., Pond, S. J. K., Varlaro, J., Stephens, K. M., & Holt, C. L. (2017). Developmental validation of the MiSeq FGx forensic genomics system for targeted next generation sequencing in forensic DNA casework and database laboratories. *Forensic Science International: Genetics, 28*, 52–70. https://doi.org/10.1016/j.fsigen.2017.01.011

Kampmann, M.-L., Buchard, A., Børsting, C., & Morling, N. (2016). High-throughput sequencing of forensic genetic samples using punches of FTA cards with buccal swabs. *Biotechniques, 61*(3), 149–151. https://doi.org/10.2144/000114453

Kidd, K. (2012). Expanding data and resources for forensic use of SNPs in individual identification. *Forensic Science International: Genetics, 6*(5), 646–652.

Klavens, A., Kollmann, D. D., Elkins, K. M., & Zeller, C. B. (2021). Comparison of DNA yield and STR profiles from the diaphysis, mid-diaphysis, and metaphysis regions of femur and tibia long bones. *Journal of Forensic Sciences, 66*(3), 1104–1113. https://doi.org/10.1111/1556-4029.14657

Kosoy, R., Nassir, R., & Tian, C. (2009). Ancestry informative marker sets for determining continental origin and admixtures proportions in common populations in America. *Human Mutation, 30*(1), 69–78.

Krenke, B. E., Nassif, N., Sprecher, C. J., Knox, C., Schwandt, M., & Storts, D. R. (2008). Developmental validation of a real-time PCR assay for the simultaneous quantification of total human and male DNA. *Forensic Science International: Genetics, 3*(1), 14–21. https://doi.org/10.1016/j.fsigen.2008.07.004

Laurent, F. X., Ausset, L., Clot, M., Jullien, S., Chantrel, Y., Hollard, C., & Pene, L. (2017). Automation of library preparation using illumina ForenSeq kit for routine sequencing of casework samples. *Forensic Science International: Genetics Supplement Series, 6*, e415–e417. https://doi.org/10.1016/j.fsigss.2017.09.156

Loreille, O. M., Diegoli, T. M., Irwin, J. A., Coble, M. D., & Parsons, T. J. (2007). High efficiency DNA extraction from bone by total demineralization. *Forensic Science International, 1*, 191–195.

Loreille, O. M., Parr, R. L., McGregor, K. A., Fitzpatrick, C. M., Lyons, C., Yang, D. Y., Speller, C. F., & Grimm, M. R. (2010). Integrated DNA and fingerprint analyses in the identification of 60-year old mummified human remains discovered in an Alaskan Glacier. *Journal of Forest Science, 55*, 813–818.

Marshall, C., Sturk-Andreaggi, K., Daniels-Higginbotham, J., Oliver, R. S., Barritt-Ross, S., & McMahon, T. P. (2017). Performance evaluation of a mitogenome capture and illumina sequencing protocol using non-probative, case-type skeletal samples: Implications for the use of a positive control in a next-generation sequencing procedure. *Forensic Science International: Genetics, 31*, 198–206.

Morrison, J., McColl, S., Louhelainen, J., Sheppard, K., May, A., Girdland-Flink, L., Watts, G., & Dawnay, N. (2020). Assessing the performance of quantity and quality metrics using the QIAGEN Investigator® Quantiplex® Pro RGQ Kit. *Science & Justice, 60*(4), 388–397. https://doi.org/10.1016/j.scijus.2020.03.002

Muller, P., Alonso, A., Barrio, P. A., Berger, B., Bodner, M., Martin, P., & Parsons, W. (2018). The DNA-SEQEX consortium. Systematic evaluation of the early access applied biosystems precision ID globalfiler mixture ID and blobalfiler NGS STR panels for the ion S5 system. *Forensic Science International: Genetics, 36*, 95–103.

Naue, J., Sänger, T., & Lutz-Bonengel, S. (2020). Get it off, but keep it: Efficient cleaning of hair shafts with parallel DNA extraction of the surface stain. *Forensic Science International: Genetics, 45*, 102210. https://doi.org/10.1016/j.fsigen.2019.102210

Niederstatter, H., Kochl, S., Grubwieser, P., Pavlic, M., Steinlechner, M., & Parson, W. (2007). A modular real-time PCR concept for determining the quantity and quality of human nuclear and mitochondrial DNA. *Forensic Science International: Genetics, 1*, 29–34.

Pajnič, Z., Obal, M. I., & Zupanc, T. (2020). Identifying victims of the largest second World war family massacre in Slovenia. *Forensic Science International, 306*, 110056. https://doi.org/10.1016/j.forsciint.2019.110056

Pakstis, A. J., Speed, W. C., Fang, R., Hyland, F. C. L., Furtado, M. H., Kidd, J. R., & Kidd, K. K. (2010). SNPs for a universal individual identification panel. *Human Genetics, 127*, 315–324.

Preuner, S., Danzer, M., Pröll, J., Pötschger, U., Lawitschka, A., Gabriel, C., & Lion, T. (2014). High-quality DNA from fingernails for genetic analysis. *Journal of Molecular Diagnostics, 16*(4), 459–466. https://doi.org/10.1016/j.jmoldx.2014.02.004

Sidstedt, M., Rådström, P., & Hedman, J. (2020). PCR inhibition in QPCR, DPCR and MPS—Mechanisms and solutions. *Analytical and Bioanalytical Chemistry, 412*(9), 2009–2023. https://doi.org/10.1007/s00216-020-02490-2

Tao, R., Wenjie, Q., Chen, C., Zhang, J., Yang, Z., Song, W., Zhang, S., & Li, C. (2019). Pilot study for forensic evaluations of the precision ID GlobalFiler™ NGS STR panel with the ion S5™ system. *Forensic Science International, 43*, 102147.

Tasker, E., LaRue, B., Beherec, C., Gangitano, D., & Hughes-Stamm, S. (2017). Analysis of DNA from post-blast pipe bomb fragments for identification and determination of ancestry. *Forensic Science International: Genetics, 28*, 195–202. https://doi.org/10.1016/j.fsigen.2017.02.016

Van Djik, E. L., Auger, H., Jaszczyszyn, & Thermes, C. (2014). Ten years of next-generation sequencing technology. *Trends in Genetics, 30*(9), 418–426.

Van Neste, C., Vandewoestyne, M., Van Criekinge, W., Deforce, D., & Nieuwerburghm, F. V. (2014). My-forensic-loci-queries (MyFLq) framework for analysis of forensic STR data generated by massive parallel sequencing. *Forensic Science International, 9*, 1–8.

Vraneš, M., Scherer, M., & Elliott, K. (2017). Development and validation of the Investigator® Quantiplex Pro Kit for QPCR-based examination of the quantity and quality of human DNA in forensic samples. *Forensic Science International: Genetics Supplement Series, 6*, e518–e519. https://doi.org/10.1016/j.fsigss.2017.09.207

Walsh, P. S., Metzger, D. A., & Higuchi, R. (1991). Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. *Biotechniques, 10*(4), 506–513.

Wang, Z., Zhou, D., Wang, H., Jia, Z., Liu, J., Qian, X., Li, C., & Hou, Y. (2017). Massively parallel sequencing of 32 forensic markers using the precision ID global filer NGS STR panel and the ion PGM system. *Forensic Science International: Genetics, 31*, 126–134.

Weber-Lehmann, J., Schilling, E., Gradl, G., Richter, D. C., Wiehler, J., & Rolf, B. (2014). Finding the needle in the haystack: Differentiating "identical" twins in paternity testing and forensics by ultra-deep next generation sequencing. *Forensic Science International, 9*, 42–46.

Xavier, C., & Parson, W. (2017). Evaluation of the Illumina ForenSeq™ DNA Signature Prep Kit-MPS forensic application for the MiSeq FGX benchtop sequencer. *Forensic Science International*, 188–194.

Yang, Y., Xie, B., & Yan, J. (2014). Application of next-generation sequencing technology in forensic science. *Genomics, Proteomics & Bioinformatics, 12*(5), 190–197. https://doi.org/10.1016/j.gpb.2014.09.001

Zeng, X., Elwick, K., Mayes, C., Takahashi, M., King, J. L., Gangitano, D., Budowle, B., & Hughes-Stamm, S. (2019). Assessment of impact of DNA extraction methods on analysis of human remain samples on massively parallel sequencing success. *International Journal of Legal Medicine, 133*(1), 51–58. https://doi.org/10.1007/s00414-018-1955-9

CHAPTER 5

Commercial kits commonly used for NGS based forensic DNA analysis

Tugba Unsal Sapan
Üsküdar University, Faculty of Engineering and Natural Science, Department of Forensic Sciences, Institute of Addiction and Forensic Sciences, Istanbul, Turkey

Introduction

DNA sequencing technology is used for amplified DNA regions in forensic genetics. The use of DNA markers for forensic purposes and their sequencing started with the discovery of restriction length polymorphism in the early 1980s and then gained momentum with short tandem repeats (STRs). In the following years, additional markers such as mitochondrial DNA (mtDNA) sequencing were developed and used in identification studies for forensic DNA analysis, especially for ancestry analysis.

In addition, since STRs spread over long intervals on the DNA sequence, they were insufficient in concluding some forensic events since they required a higher amount of DNA, and to elucidate these events, single-nucleotide polymorphisms (SNPs), which were located at fewer short intervals in the DNA sequence and allowed to work with less DNA amount started to be used.

Since SNPs are biallelic, their discriminatory power is low alone, and it is necessary to study 50–60 SNPs to reach the discrimination power of 13 STRs. In this way, as a result of the limitations of Sanger sequencing technology in a multilocus study, such as not being able to analyze in one run, researchers have turned to next-generation sequencing (NGS) technologies as an alternative in forensic DNA analysis. In forensic DNA analysis, NGS technology has provided significant advantages for SNP or mtDNA analysis of gradient samples, the analysis of challenging paternity cases, and new species analysis such as ancestry and phenotype determination. In addition, it has become the preferred system in recent years, allowing many systems to be analyzed in a single workflow (Børsting & Morling, 2015; Yang et al., 2014).

Development of next-generation sequencing systems

NGS technology refers to nonSanger sequencing based on high-throughput DNA sequencing technology. With this technology, millions or billions of DNA molecules can be sequenced in parallel. Thus, the throughput increases significantly and minimizes the need for the fragment analysis method frequently used in Sanger sequencing. Second-

Next Generation Sequencing (NGS) Technology in DNA Analysis
ISBN 978-0-323-99144-5, https://doi.org/10.1016/B978-0-323-99144-5.00018-4
© 2024 Elsevier Inc.
All rights reserved.

generation sequencing technology based on loop sequence sequencing, which can analyze large numbers of samples simultaneously, has also been developed, as well as third-generation sequencing technologies that can determine the basic composition of single DNA molecules (van Dijk et al., 2014).

Second-generation NGS systems

Different methods are used in the sequencing of NGS systems:

- Pyrosequencing: Using the emulsion PCR technique with oligonucleotides attached to clonal amplification beads.
- Sequencing by synthesis: It is the synthesis technology that follows the addition of labeled nucleotides as the DNA strand is copied.
- Sequencing by ligation: It is a DNA sequencing method that uses the DNA ligase enzyme to identify nucleotides at a specific position in the DNA sequence.
- Semiconductor sequencing: This is a DNA sequencing method based on the detection of hydrogen ions released during the polymerization of DNA exposed to hydrogen ion bombardment.
- Shotgun method: used for sequencing large genomes. The method involves breaking the genome into a collection of small DNA fragments that are sequenced individually. A computer program looks for overlaps in the DNA sequences and uses them to place the individual fragments in their correct order to reconstitute the genome. This method is the most common of the NGS methods. This method is not used in forensic analyses because it works with large genomes and is not suitable for pairwise comparison.
- Targetted (re) sequencing method: This is another method used. It provides amplification of selected regions by PCR or sequencing of the desired region using probes or both probes and enzymes. This method allows more samples to be run simultaneously. Unlike the Shotgun method, Targeted (re) sequencing is suitable for use in forensic sciences. With this method, identification, phenotype predictions, and tissue typing can be done (Alvarez-Cubero et al., 2017; Børsting & Morling, 2015; Daniel et al., 2015; Heather & Chain, 2016; Mamanova et al., 2010; Mardis, 2008).

The first development in the field of NGS was made by 454 Corporation in 2005. The company developed the pyrosequencing method using the emulsion PCR technique with oligonucleotides attached to clonal amplification beads and named its first system "Genome Sequencer 20". With this device, the human genome has been sequenced in 5 months. However, as a result of the development of technology every second and the development of new methods with new systems, this device is not produced today. This system was purchased by Roche in the following years and updated with the "Roche GS FLX Titanium" platform. The 454 Genome Sequencer has approximately 2 million wells for analysis, each in the form of a 28-μm-diameter bead coated with

single-stranded copies of the target sequence. Current 454/Roche GS FLX Titanium platforms can sequence DNA regions of 300–500 nt in length in approximately 1.5 million wells in a single analysis (Kircher & Kelso, 2010; McGinn & Glynne Gut, 2013).

The second NGS system was created by Solexa in 2006. It launched the "Genome Analyzer" system. Solexa merged with Illumina in 2007 and introduced the "Illumina HiSeq 2500" system. With Illumina HiSeq 2500, the human genome was sequenced 30 times more in 1 day. In 2011, Illumina introduced systems named iSeq 100 System, Miniseq System, MiSeq System, NextSeq 550, and NextSeq 1000–2000, respectively, which enable lower volumes of data to be studied in a shorter time. The technique used in these systems is the SNaPshot sequencing (sequencing by synthesis) method. Up to 400 million reads can be made for target DNA sequencing of 150–300 nt per analysis in the wells of Illumina systems (Alvarez-Cubero et al., 2017; McGinn & Glynne Gut, 2013; Nextseq 550 System, Specification Sheet, 2021).

The third NGS system is owned by Applied Biosystems. In 2007, Applied Biosystems (ABI) released the "SOLiD" second-generation sequencing system based on the oligonucleotide ligation technique and binary coding system. In 2010, they developed the ABI "Ion Torrent" system, a faster and lower-cost sequencer based on semiconductor technology. The Ion Torrent system uses fluorescence, chemiluminescence, or enzyme-cascading techniques to sequence signal detection. In the following years, Thermo Fisher Scientific, which acquired Applied Biosystems, developed the NGS technologies Ion Torrent, Personal Genome Machine (PGM), Ion Proton, and Ion S5 sequencers, respectively. The latest technology Ion S5 system chips (well systems) can perform 2 million–130 million reads per analysis of 200 nt target DNA sequencing (Ion S5 and Ion S5 XL Systems Specification Sheet, 2017; Kircher & Kelso, 2010; Meiklejohn & Robertson, 2017).

Third-generation sequencing systems

In addition to these methods, new technologies called third-generation sequencing have begun to develop. Real-time reading is possible with third-generation sequencing. It is faster than the systems of second-generation technology and allows the study of long DNA fragments without the need for amplification in PCR. One of the third-generation sequencing technologies is the "MinION" instrument produced by Oxford Nanopore Technologies. MinION is a technique developed in 2014. Nanopores are engineered proteins. A voltage is applied in solution to nanopores made of synthetic polymers embedded in an electrically resistive membrane. Characteristic changes occur as a result of the current that emerges while applying voltage. In this way, nucleotides are identified (Peterson et al., 2019; Wilson et al., 2019).

Apart from the MinION device, Oxford Nanopore Technologies has produced another device called "GridION". The GridION system detects a large number of

nanopores that can be measured and analyzed in real time. MinIONTM is a disposable version of GridION (Alvarez-Cubero et al., 2017; McGinn & Glynne Gut, 2013).

Use of NGS in forensic sciences

The application of DNA technologies in forensic research has made DNA analysis an important tool in forensic science. Compared to other fields of life sciences, forensic DNA analysis is an area that needs analysis with high accuracy and reproducibility as it has a low copy number and highly degraded and contaminated samples from the crime scene (Alvarez-Cubero et al., 2017; Elkins et al., 2021).

The Sanger methodology has been recognized as the gold standard for DNA sequencing. However, CE-based analysis also has disadvantages. The inability to analyze multiple genetic polymorphisms in a single reaction using a single workflow, low-resolution genotyping of available markers, and the inability to obtain useful genomic information in degraded DNA samples are examples of disadvantages (Elkins et al., 2021).

Analyses are carried out with various types of genetic markers for identification in forensic laboratories (autosomal-STR, Y-STR, X-STR, InDel, mtDNA, autosomal SNP, X-SNP, and Y-SNP, ancestry informative markers: Ancestry SNPs and phenotypic SNPs). Most of these markers can be analyzed with the classical Sanger sequencing method in 1 day, while some may take a day or two. However, two different types of analysis cannot be done in a single workflow or in a single reaction. For this reason, commercial kits, which offer the opportunity to work with these markers as an alternative to suitable Sanger sequencing systems in NGS technologies, have been developed by companies and started to be used routinely in forensic laboratories (Elkins et al., 2021).

Commercial NGS kits used for forensic analysis

There are NGS panels used in the field of forensic sciences that are routinely used by two companies: Thermo Fisher Scientific and the Illumina-Verogen partnership.

Illumina-Verogen partnership

The system developed by Illumina in cooperation with the Forensic Sciences community with the modifications they made to MiSeq, one of the NGS systems, is known as the "Illumina MiSeq FGx Forensic Genomics System". This system is completely designed for forensic analysis and provides direct results from the DNA sample. Compatible with the Illumina MiSeq FGx Forensic Genomics System, Verogen has also developed the following five kits used in forensic science (Jäger et al., 2017).

- ForenSeq DNA Signature Prep kit.
- ForenSeq Kintelligence kit.
- ForenSeq MainstAY kit.
- ForenSeq mtDNA control region kit.

- ForenSeq mtDNA whole genome kit.

ForenSeq DNA Signature Prep Kit

The ForenSeq DNA Signature Prep Kit is a NGS-based kit designed to be uploaded to the National DNA Index System for forensic cases. It contains more than 200 markers together with STRs and SNPs: Amelogenin, 27 autosomal STRs, 7 X chromosomal STRs, 24 Y chromosomal STRs (Y-STRs), 94 identity-informative SNPs, 22 phenotypic-informative SNPs, and 56 ancestry-informative SNPs. It is a kit that combines all markers into one streamlined workflow (Guo et al., 2017).

ForenSeq Kintelligence Kit

Forensic genetic genealogy (FGG) analysis provides important contributions within the scope of cold cases, missing persons, and innocence projects in forensic sciences. FCG analysis combines microarray genotyping and whole-genome sequencing to create DNA profiles. It compares the DNA profile in this database using databases such as GEDmatch, which is considered the largest database. It determines whether there are genetic relatives of the perpetrators of unsolved events or missing persons among those who voluntarily register their DNA information in the database.

The ForenSeq Kintelligence Kit, which sequences 10,230 SNPs, makes it possible to compare the target DNA with the DNA of the individuals in the database. The ForenSeq Kintelligence Kit does not analyze disease-related DNA regions. It also includes the content of the ForenSeq DNA Signature Prep Kit (identity markers, biogeographic ancestry, phenotype determination) as it is used for forensic purposes. It includes 56 Ancestry SNPs, 94 Identity SNPs, 9867 Kinship SNPs, 22 Phenotype SNPs, 106 X-SNPs, and 85 Y-SNPs (ForenSeq Kintelligence Kit, Datasheet, 2021).

ForenSeq MainstAY Kit

The ForenSeq MainstAY Kit is a panel that enables profiles to be obtained by analyzing 25 autosomal STRs (Au-STR) and 27 Y-STRs together in MiSeq FGx Sequencing Systems. It aims to reduce the workload and cost in forensic laboratories by combining two separate workflows with a single amplification and sequencing.

STR loci included in the ForenSeq MainstAY Kit:

AuSTRs: D1S1656, vWA, TPOX, D12S391, D2S1338, Penta E, D3S1358, D16S539, D4S2408, D17S1301, FGA, D18S51, D5S818, D19S433, CSF1PO, D20S482, D6S1043, D21S11, Penta D, D7S820, D8S1179, D22S1045, D9S1122, TH01, D10S1248, amelogenin.

Y-STRs: DYF387S1, DYS439, DYS19, DYS448, D2S441, D13S317, DYS385a-b, DYS460, DYS389I, DYS481, DYS389II, DYS505, DYS390, DYS522, DYS391, DYS533, DYS392, DYS549, DYS393, DYS570, DYS437, DYS576, DYS438, DYS612, DYS635, Y-GATA-H4, DYS643 (ForenSeq MainstAY Kit, Datasheet, 2021).

ForenSeq mtDNA Control Region Kit

The ForenSeq mtDNA Control Region Kit is a product that enables the analysis of the control region of mtDNA with a low amount of DNA, especially suitable for analysis in degraded or trace amounts of biological samples. To improve performance on highly degraded samples, this kit includes 122 primers of 18 amplicons <150 bp in length, recently designed by evaluating mtDNA variant and frequency data (Data Sheet: ForenSeq mtDNA Control Region, 2019; Holt et al., 2021).

ForenSeq mtDNA Whole Genome Kit

The ForenSeq mt Whole Genome Kit is an NGS-based product containing 663 primers from 245 amplicons for sequencing the entire mtDNA in challenging samples. The kit allows for cross-integrated diversity for genome analyses. Limiting the kit to a low amount of DNA (100 pg) may allow an error-free analysis of the 16,569 bp genome (ForenSeq mtDNA Whole Genome Kit, Datasheet, 2020; Holt et al., 2021).

Thermo fisher scientific

The following 9 kits, which are compatible with the company's Ion Torrent PGM System and HID Ion GeneStudio S5 System devices, are used in forensic sciences.

- Ion AmpliSeq DNA phenotyping panel.
- Ion ampliseq phenotrivium panel.
- Precision ID mtDNA control region panel.
- Precision ID mtDNA whole genome.
- Precision ID ancestry panel.
- Precision ID identity panel.
- Precision ID globalFiler NGS STR v2.
- Ion ampliSeq HID Y-SNP research panel v1.
- Ion ampliSeq MH-74 plex research panel.
- Ion ampliSeq VISAGE-basic tool research panel.

Ion AmpliSeq DNA Phenotyping Panel

DNA phenotyping can be performed to aid the investigation in cases where there is no match in the DNA database of STR profiles. The Ion AmpliSeq DNA Phenotyping Panel is used to predict hair and eye color using 24 phenotype SNPs from the HIrisPlex system. This panel includes 23 SNPs and an indel from the following genes: MC1R, SLC45A2, ASIP/PIGU, EXOC2, HERC2, IRF4, KITLG, OCA2, TYR, SLC24A4, and TRYP1. In combination with black, brown, red, or blond hair, it provides results showing the probability of a blue, intermediate, or brown eye color (Walsh et al., 2013).

Ion AmpliSeq Pheno Trivium Panel

In the field of forensic genetics, studies are continuing to use alternative polymorphic regions to achieve success with challenging samples. In particular, the use of biogeographic ancestry and DNA phenotypes can be helpful in investigations in cases where there are unsolved cases or no suspects. The Ion AmpliSeq PhenoTrivium Panel, containing a total of 200 autosomal SNPs and 120 Y chromosomal SNPs in a single panel, includes biogeographic ancestry, DNA phenotyping, and male lineage markers. With this kit, it is possible to determine lineage and phenotype with 125 pg of genomic DNA (gDNA) (Diepenbroek et al., 2020).

Precision ID mtDNA control region panel

This is an NGS method for mtDNA analysis developed for forensic applications. mtDNA analysis is an important method for DNA analysis when the nuclear DNA analyzed is insufficient when samples such as hair, teeth, and bone that do not contain follicles are taken from crime scenes, missing person cases, or DVI. It is a multiplex method of 2 pools targeting the 1.2 kb (kilobase) control region of the mitochondrial genome, which includes HV-I, HV-II, and HV-III. It has been produced as a new sequencing method to obtain optimal results from samples that are likely to degrade, such as hair, teeth, and bones. This panel is based on Ion AmpliSeq technology (Precision ID mtDNA Panel Analysis User Guide, 2016).

Precision ID mtDNA whole genome panel

This is a 2-pool multiplex method targeting the entire mitochondrial genome, i.e., 16,569 bp. It contains 81 primer pairs with minimal primer overlap between pools. It has been produced as a new sequencing method to obtain optimum results from samples that are likely to degrade, such as hair, teeth, and bones (Precision ID mtDNA Panel Analysis User Guide, 2016).

Precision ID ancestry panel

This panel contains 165 SNPs that provide biogeographic and genealogical information. Of the 165 SNP sites, 55 were derived from publications by Kenneth Kidd. The remaining 123 SNP sites were obtained from publications by Michael Seldin. Generally, the amplicons of degraded samples are less than 130 bp. However, with this panel, degraded samples can be studied (Kidd, 2015, pp. 1–80; Pakstis et al., 2010).

Precision ID identity panel

This panel is one of the first NGS systems developed for identification and uniquely identifies samples that have degraded or are in good condition. It is a panel with high

discrimination power using 34 upper Y-Clade SNPs and 90 autosomal SNPs, for a total of 124 with high heterozygosity. It is used for degraded or challenging forensic samples. This kit's discrimination of individuals is similar to STR genotype match probabilities used in forensic sciences (between 1×10^{-31} and 6×10^{-35}) (Karafet et al., 2008; Kidd, 2015, pp. 1–80; Pakstis et al., 2010; Phillips et al., 2007).

Precision ID Globalfiler NGS panel v2

This panel includes a total of 35 markers, 20 CODIS core STR regions, 11 nonCODIS STRs, and 4 gender regions to aid in the interpretation of challenging case examples.

20 CODIS STR regions: TPOX, D3S1358, FGA, D5S818, CSF1PO, D7S820, D8S1179, TH01, vWA, D13S317, D16S539, D18S51, D21S11, D1S1656, D2S1338, D2S441, D10S1248, D12S391, D19S433, and D22S1045;

11 nonCODIS STR regions: D1S1677, D2S1776, D3S4529, D4S2408, D5S2800, D6S1043, D6S474, D12ATA63, D14S1434, Penta E, Penta D, and *4 sex regions*: Amelogenin, DYS391, SRY, and Y-indel (Ragazzo et al., 2020; Tao et al., 2019).

Ion AmpliSeq HID Y-SNP research panel v1

It is an NGS-based kit developed for ancestry and paternal biogeographic ancestry using Y chromosomal SNPs and Y chromosomal haplogroups. This panel allows the analysis of 859 phylogenetic Y-SNPs, 640 of which are Y haplogroups (Arwin et al., 2019; Claerhout et al., 2021).

Ion AmpliSeq MH-74 plex research panel

It is a 157–325 bp panel consisting of a total of 230 SNPs containing 74 more microhaplotypes in addition to 130 microhaplotypes previously optimized by the Kidd lab. Microhaplotypes are particularly suitable for the analysis of mixed samples among forensic samples and also contain SNPs that determine biogeographical ancestry. It is a panel containing alternative polymorphic regions for cases where conventional STR capillary electrophoresis typing fails. It is possible to obtain a full profile with 50 pg of DNA (Oldoni et al., 2020).

Ion AmpliSeq VISAGE-basic tool research panel

In forensic science, DNA phenotyping is done to aid the investigation when there is no DNA database or no match in the database. The investigation can be assisted by finding data on a person's phenotype, including appearance characteristics and biogeographic ancestry.

The Ion AmpliSeq VISible Attributes through GEnomics Basic Tool Research Panel includes 41 SNPs of the HIrisPlex-S assay for phenotype and biogeographic ancestry, with an additional 115 biogeographic ancestry markers capable of detecting population

differentiation from 7 continents, for a total of 153 markers. A full profile can be obtained with 100 pg of gDNA with this kit (Arwin et al., 2019; Xavier et al., 2020).

Conclusion

NGS technologies, also known as massively parallel sequencing (MPS), have been used in genetic research since the 2000s. Various methods have been developed that are faster, more efficient, and less costly for NGS. As NGS methods and platforms have developed, the quality of sequences has increased, and in recent years, forensic genetics laboratories have turned to this technology. NGS technology allows the analysis of combining many systems (STRs, SNPs, and mRNA) at the same time. In addition, with this technology, DNA database creation, ancestry and phenotype determination, identification of identical twins in the forensic field, distinction, identification of body fluids, and species, and microbiology analysis can be performed.

NGS for forensic genetic analysis has many advantages over the traditional CE method, although the process of transition to laboratories has been difficult and there have been some difficulties in its applications. It has also not completely replaced conventional CE technology and kits. Its difficulties include the inability to create validated and powerful kits, high equipment costs, limited NGS training of analysts, validation, and optimization studies.

However, in recent years, both researchers and commercial companies have accelerated their studies on this subject, allowing NGS to become an established system in the field of forensic sciences. The development of commercial panels not only enabled the use of an alternative method to the CE system in forensic genetic analysis but also gave direction to investigations with forensic analyses at a point where CE technology was insufficient. Owing to their advantages, such as the application of forensic genetic studies of more than one technique in a single analysis, they will appear as systems used in a large part of forensic laboratories, shortly haplogroups next-generation to type.

References

Alvarez-Cubero, M. J., Saiz, M., Martínez-García, B., Sayalero, S. M., Entrala, C., Lorente, J. A., & Martinez-Gonzalez, L. J. (2017). Next generation sequencing: An application in forensic sciences? *Annals of Human Biology, 44*(7), 581–592. https://doi.org/10.1080/03014460.2017.1375155

Arwin, R., Mannis, V. O., Diego, M. G., Peter, D. K., Kees, V. D., Sharon, W., Lagace, R., & Kayser, M. (2019). Forensic Y-SNP analysis beyond SNaPshot: High-resolution Y-chromosomal haplogrouping from low quality and quantity DNA using Ion AmpliSeq and targeted massively parallel sequencing. *Forensic Science International: Genetics, 41*, 93–106. https://doi.org/10.1016/j.fsigen.2019.04.001

Børsting, C., & Morling, N. (2015). Next generation sequencing and its applications in forensic genetics. *Forensic Science International: Genetics, 18*, 78–89. https://doi.org/10.1016/j.fsigen.2015.02.002

Claerhout, S., Verstraete, P., Warnez, L., Vanpaemel, S., Larmuseau, M., Decorte, R., & Gojobori, T. (2021). CSYseq: The first Y-chromosome sequencing tool typing a large number of Y-SNPs and Y-

STRs to unravel worldwide human population genetics. *PLoS Genetics, 17*(9), e1009758. https://doi.org/10.1371/journal.pgen.1009758

Daniel, R., Santos, C., Phillips, C., Fondevila, M., van Oorschot, R. A. H., Carracedo, Á., Lareu, M. V., & McNevin, D. (2015). A SNaPshot of next generation sequencing for forensic SNP analysis. *Forensic Science International: Genetics, 14*, 50–60. https://doi.org/10.1016/j.fsigen.2014.08.013

Data Sheet: ForenSeq mtDNA Control Region. (2019). *Verogen inc.* https://verogen.com/wp-content/uploads/2020/08/191115-ForenSeq-mtDNA-Control-Region-datasheet.pdf.

Diepenbroek, M., Bayer, B., Schwender, K., Schiller, R., Lim, J., Lagacé, R., & Anslinger, K. (2020). Evaluation of the ion AmpliSeq phenotrivium panel: MPS-based assay for ancestry and phenotype predictions challenged by casework samples. *Genes, 11*(12), 1–24. https://doi.org/10.3390/genes11121398

van Dijk, E. L., Auger, H., Jaszczyszyn, Y., & Thermes, C. (2014). Ten years of next-generation sequencing technology. *Trends in Genetics, 30*(9), 418–426. https://doi.org/10.1016/j.tig.2014.07.001

Elkins, K. M., Berry, H. E., & Reese, K. R. (2021). Applications of NGS in DNA analysis. In H. R. Dash, P. Shrivastava, & J. A. Lorente (Eds.), *Handbook of DNA profiling* (pp. 1–18). Singapore: Springer. https://doi.org/10.1007/978-981-15-9364-2_19-1

ForenSeq Kintelligence Kit, Datasheet. (2021). *Verogen inc.* https://verogen.com/wp-content/uploads/2021/03/forenseq-kintelligence-datasheet-vd2020054-a.pdf.

ForenSeq MainstAY Kit, Datasheet. (2021). *Verogen inc* (pp. 1–7). https://verogen.com/wp-content/uploads/2021/07/ForenSeq-MainstAY-Kit-Datasheet-Document-VD2020055.pdf.

ForenSeq mtDNA Whole Genome Kit, Datasheet. (2020). *Verogen inc.* https://verogen.com/wp-content/uploads/2020/08/forenseq-mtdna-whole-genome-data-sheet-vd2020011.pdf.

Guo, F., Yu, J., Zhang, L., & Li, J. (2017). Massively parallel sequencing of forensic STRs and SNPs using the Illumina® ForenSeq DNA signature Prep kit on the MiSeq FGx forensic genomics system. *Forensic Science International: Genetics, 31*, 135–148. https://doi.org/10.1016/j.fsigen.2017.09.003

Heather, J. M., & Chain, B. (2016). The sequence of sequencers: The history of sequencing DNA. *Genomics, 107*(1), 1–8. https://doi.org/10.1016/j.ygeno.2015.11.003

Holt, C. L., Stephens, K. M., Walichiewicz, P., Fleming, K. D., Forouzmand, E., & Wu, S. F. (2021). Human mitochondrial control region and mtgenome: Design and forensic validation of NGS multiplexes, sequencing and analytical software. *Genes, 12*(4). https://doi.org/10.3390/genes12040599

Ion. (2017). S5 and ion S5 XL systems specification Sheet (COL31745 0617). *Thermo fischer scientific inc.; ion GeneStudio S5 series speed and throughput flexibility across*. https://tools.thermofisher.com/content/sfs/brochures/Ion-S5-S5XL-SpecSheet.pdf.

Jäger, A. C., Alvarez, M. L., Davis, C. P., Guzmán, E., Han, Y., Way, L., Walichiewicz, P., Silva, D., Pham, N., Caves, G., Bruand, J., Schlesinger, F., Pond, S. J. K., Varlaro, J., Stephens, K. M., & Holt, C. L. (2017). Developmental validation of the MiSeq FGx forensic genomics system for targeted next generation sequencing in forensic DNA casework and database laboratories. *Forensic Science International: Genetics, 28*, 52–70. https://doi.org/10.1016/j.fsigen.2017.01.011

Karafet, T. M., Mendez, F. L., Meilerman, M. B., Underhill, P. A., Zegura, S. L., & Hammer, M. F. (2008). New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. *Genome Research, 18*(5), 830–838. https://doi.org/10.1101/gr.7172008

Kidd, K. K. (2015). *Further development of SNP panels for forensics, NIJ final technical report* (pp. 1–80). U.S. Department of Justice. www.ojp.gov/pdffiles1/nij/grants/249548.pdf.

Kircher, M., & Kelso, J. (2010). High-throughput DNA sequencing–concepts and limitations. *BioEssays, 32*(6), 524–536. https://doi.org/10.1002/bies.200900181

Mamanova, L., Coffey, A. J., Scott, C. E., Kozarewa, I., Turner, E. H., Kumar, A., Howard, E., Shendure, J., & Turner, D. J. (2010). Target-enrichment strategies for next-generation sequencing. *Nature Methods, 7*(2), 111–118. https://doi.org/10.1038/nmeth.1419

Mardis, E. R. (2008). Next-generation DNA sequencing methods. *Annual Review of Genomics and Human Genetics, 9*, 387–402. https://doi.org/10.1146/annurev.genom.9.081307.164359

McGinn, S., & Glynne Gut, I. (2013). DNA Sequencing - spanning the generations. *New Biotechnology, 30*(4), 366–372. https://doi.org/10.1016/j.nbt.2012.11.012

Meiklejohn, K. A., & Robertson, J. M. (2017). Evaluation of the precision ID identity panel for the ion torrent ™ PGM ™ sequencer. *Forensic Science International: Genetics, 31*, 48–56.

NextSeqTM 550 System. (2021). *Specification sheet* (770th-2013th-053-H. QB11101. ed., pp. 1–7). Illumina, Inc.

Oldoni, F., Bader, D., Fantinato, C., Wootton, S. C., Lagacé, R., Kidd, K. K., & Podini, D. (2020). A sequence-based 74plex microhaplotype assay for analysis of forensic DNA mixtures. *Forensic Science International: Genetics, 49*. https://doi.org/10.1016/j.fsigen.2020.102367

Pakstis, A. J., Speed, W. C., Fang, R., Hyland, F. C. L., Furtado, M. R., Kidd, J. R., & Kidd, K. K. (2010). SNPs for a universal individual identification panel. *Human Genetics, 127*(3), 315–324. https://doi.org/10.1007/s00439-009-0771-1

Peterson, L. M., Martin, I. W., Moschetti, W. E., Kershaw, C. M., & Tsongalis, G. J. (2019). Third generation sequencing in the clincal laboratory: Exploring the advantages and challenges of nanopore sequencing. *Journal of Clinical Microbiology, 58*(1), e01315–e01319. https://doi.org/10.1128/JCM.01315-19

Phillips, C., Fang, R., Ballard, D., Fondevila, M., Harrison, C., Hyland, F., Musgrave-Brown, E., Proff, C., Ramos-Luis, E., Sobrino, B., Carracedo, A., Furtado, M. R., Court, D. S., & Schneider, P. M. (2007). Evaluation of the Genplex SNP typing system and a 49plex forensic marker panel. *Forensic Science International: Genetics, 1*(2), 180–185. https://doi.org/10.1016/j.fsigen.2007.02.007

Precision ID mtDNA panel analysis user guide. (2016). *Thermo Fisher scientific* (MAN0015910 ed.) http://file:///C:/Users/dell2/Desktop/MAN0015910_PrecisionID_mtDNA_Panel_UG.pdf.

Ragazzo, M., Carboni, S., Caputo, V., Buttini, C., Manzo, L., Errichiello, V., Puleri, G., & Giardina, E. (2020). Interpreting mixture profiles: Comparison between precision ID GlobalFiler NGS STR panel v2 and traditional methods. *Genes, 11*(6), 591. https://doi.org/10.3390/genes11060591

Tao, R., Qi, W., Chen, C., Zhang, J., Yang, Z., Song, W., Zhang, S., & Li, C. (2019). Pilot study for forensic evaluations of the Precision ID GlobalFiler^{TM} NGS STR Panel v2 with the Ion S5 system. *Forensic Science International: Genetics, 43*. https://doi.org/10.1016/j.fsigen.2019.102147

Walsh, S., Liu, F., Wollstein, A., Kovatsi, L., Ralf, A., Kosiniak-Kamysz, A., Branicki, W., & Kayser, M. (2013). The HIrisPlex system for simultaneous prediction of hair and eye colour from DNA. *Forensic Science International: Genetics, 7*(1), 98–115. https://doi.org/10.1016/j.fsigen.2012.07.005

Wilson, B. D., Eisenstein, M., & Soh, H. T. (2019). *High-fidelity nanopore sequencing of ultra-short DNA targets*. Analytical Chemistry. https://doi.org/10.1021/acs.analchem.9b00856

Xavier, C., de la Puente, M., Mosquera-Miguel, A., Freire-Aradas, A., Kalamara, V., Vidaki, A., Gross, T. E., Revoir, A., Pośpiech, E., Kartasińska, E., Spólnicka, M., Branicki, W., Ames, C. E., Schneider, P. M., Hohoff, C., Kayser, M., Phillips, C., & Parson, W. (2020). Development and validation of the VISAGE AmpliSeq basic tool to predict appearance and ancestry from DNA. *Forensic Science International: Genetics, 48*. https://doi.org/10.1016/j.fsigen.2020.102336

Yang, Y., Xie, B., & Yan, J. (2014). Application of next-generation sequencing technology in forensic science. *Genomics, Proteomics & Bioinformatics, 12*(5), 190–197. https://doi.org/10.1016/j.gpb.2014.09.001

This page intentionally left blank

CHAPTER 6

Applications of nanopore sequencing for forensic analysis

Olivier Tytgat and Filip Van Nieuwerburgh
Ghent University, Laboratory of Pharmaceutical Biotechnology, Ghent, Belgium

Introduction

DNA sequencing for forensic purposes is currently almost exclusively performed by second-generation sequencing methods (Bruijns et al., 2018). Verogen and Thermo Fisher Scientific are among the most important providers, both relying on sequencing-by-synthesis (SBS). Verogen commercializes the MiSeq FGx Sequencing System, based on the Illumina platform. Thermo Fisher Scientific distributes the Ion Torrent systems (Churchill et al., 2016; Zhang et al., 2017). Both technologies yield high-accuracy reads at sufficient throughput. However, their widespread introduction into routine forensic use is hampered by some limitations. Firstly, SBS-based systems are short read sequencing systems with a maximal read length of about 500 nucleotides, making assay design and read alignment of long short tandem repeats (STR) loci more challenging. Secondly, the acquisition of such a technology requires a large capital investment. Third, the cost per sample is pronouncedly higher compared to capillary electrophoresis (CE). Moreover, highly trained staff is required for sample and library preparation. Lastly, analysis is performed in centralized laboratories.

The introduction of nanopore sequencing, a third-generation sequencing system, in the field of forensic DNA genotyping avoids the need for a large capital investment. Oxford Nanopore Technologies (ONT) commercializes a handheld sequencing device, the MinION, at a negligible cost (Jain et al., 2016). This long-read sequencing yields reads spanning more than 500 nucleotides. Moreover, the device's small footprint allows for sequencing in remote locations, e.g., at a crime scene. Devices capable of yielding higher throughputs than the MinION device, namely the GridION and the PromethION, are also available on the market. These devices are more suited for centralized laboratories performing routine forensic DNA analysis.

Although the general concept of nanopore sequencing dates back to an idea sketched in 1989 by David Deamer in his notebook, nanopore sequencing became available only in 2014 (Deamer et al., 2016). Nevertheless, this technology is rapidly being explored in the forensic field. In this chapter, we will discuss the nanopore sequencing technology in detail, followed by an overview of possible applications in forensic investigations. Lastly,

Next Generation Sequencing (NGS) Technology in DNA Analysis
ISBN 978-0-323-99144-5, https://doi.org/10.1016/B978-0-323-99144-5.00027-5
© 2024 Elsevier Inc.
All rights reserved.

we will elaborate on the challenges that should be overcome for nanopore sequencing to become a routine application in forensics.

Nanopore sequencing

In contrast to most other systems available on the market, nanopore sequencing does not rely on sequencing by synthesis. Instead, individual DNA molecules are sequenced in real-time without the intervention of a polymerase enzyme. The heart of this technology consists of an array of nanopores, which are small apertures embedded in an electrically resistant membrane. As shown in Fig. 6.1, this membrane separates two reservoirs filled with ionic fluid from each other. Upon applying a voltage over the membrane, a current of ions will flow through these nanopores. Nanopore sequencing is performed by threading a DNA molecule through a nanopore. A motor protein is attached to the nanopore to regulate the speed at which a DNA strand is translocated through the pore and to unwind double-stranded DNA. The presence of a DNA molecule alters the ionic current by partially blocking the aperture. As each nucleotide modifies the current in a slightly different way, the obtained current profile, called a 'squiggle', can be translated to a nucleotide sequence (Kasianowicz et al., 1996). This process is called base calling and is performed by machine learning algorithms. Advanced computational methodologies are necessary, as at each moment in time, a sequence of 5—7 nucleotides is present in

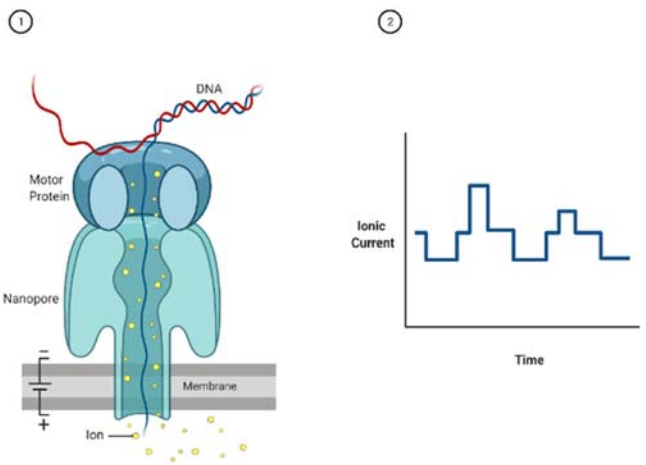


Figure 6.1 General principle of nanopore sequencing. Left: a motor protein, attached to a nanopore protein, unwinds a dsDNA strand. One of both strands is translocated through the nanopore, which is embedded in a voltage-biased membrane. The presence of a DNA strand alters the ion flow through the nanopore. Right: The resulting ionic current can be used to decode the DNA sequence, as the disruption of the ionic current is characteristic for each nucleotide passing through the pore. *(Created with BioRender.com.)*

a pore, and thus a combination of nucleotides dominates the current signal. This makes the deconvolution of a squiggle into a nucleotide sequence extremely challenging.

ONT uses engineered proteins as nanopores. In general, two types of nanopores are currently used for sequencing, i.e., the R9 and the R10 pores. For each type of pore, multiple iterations have been commercialized. As nanopore sequencing can still be considered a relatively recent technology, ONT is continuously working on improving its technology. At the time of writing, the state-of-the-art pore types were R9.4 and R10.4. The R10-series of nanopores is more recent than the R9-series and is characterized by a dual reader head, as shown in Fig. 6.2 (Van der Verren et al., 2020). Sequencing using the R9 nanopores, which only contain a single reader head, proved to be troublesome for DNA strands containing homopolymers (Eskola et al., 2017). Threading a homopolymer through the nanopore gives rise to a constant current signal over a certain time period, challenging the base calling process. The dual reader head of the R10-series was designed to improve the base calling accuracy of homopolymers as more nucleotides dominate the signal.

Besides different types of nanopores, ONT also offers a broad range of sequencing instruments, each characterized by its throughput. The smallest device is the MinION, which weighs less than 0.5 kg. It accommodates the MinION flow cells, containing an array of 2048 nanopores and the electronics needed for sequencing (Jain et al., 2016). Five hundred twelve of these nanopores can be sequenced in parallel. More recently, the Flongle adapter and Flongle flow cells were introduced. Flongle flow cells only contain 126 channels and can be used on a MinION device by using the Flongle adapter. The electronics needed for sequencing are embedded in this reusable adapter, lowering the cost significantly. MinION and Flongle flow cells are also compatible with the GridION device, a benchtop device capable of running 5 flow

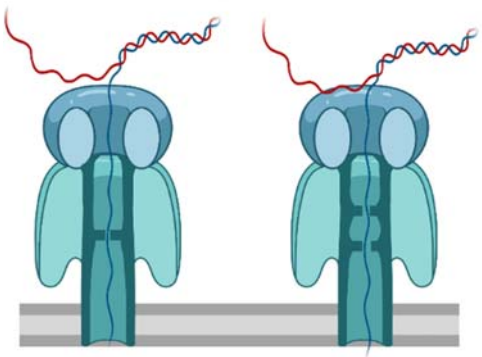


Figure 6.2 Comparison of R9 and R10 nanopores. An R9 nanopore is characterized by a single reader head, which is, in essence, a constriction. As the R10 nanopores have a dual reader head, more nucleotides dominate the signal. *(Created with BioRender.com.)*

cells in parallel. For projects that require high throughputs, a PromethION device and compatible flow cells are most suited. One PromethION can run 24 (PromethION 24) or 48 (PromethION 48) flow cells in parallel, each capable of using 2675 nanopores in parallel.

Forensic applications of nanopore sequencing

Short tandem repeat genotyping

STRs are the most commonly used marker in forensic genetics (Butler, 2005). The advantages of sequencing STRs instead of performing length-based analysis using CE are well known (Bruijns et al., 2018; Børsting & Morling, 2015). MPS yields a more informative output, as besides length polymorphisms, sequence polymorphisms can also be detected, e.g., single nucleotide polymorphisms (SNPs) in the repeat region and the surrounding regions. As a consequence, multiple research groups are investigating the performance of STR genotyping by nanopore sequencing. Unfortunately, due to their highly repetitive nature, STRs are challenging targets for sequencing technologies. Moreover, the read quality of nanopore reads is still lower compared to, e.g., Illumina sequencing (Rang et al., 2018). The first report on nanopore sequencing of forensic STRs, published by Cornelis et al. demonstrated those issues (Cornelis et al., 2018). After performing a 14-locus multiplex PCR, the resulting amplicons were randomly concatenated. This additional step was required as the minimal read length at the time was 100 nucleotides, which exceeded the length of some amplicons. Sequencing was performed using an R7 flow cell, which is no longer available. Only partial forensic profiles could be obtained due to a high sequencing error rate. Moreover, a particularly high error rate was observed in homopolymeric regions, mostly partial deletions. These findings were corroborated by Asogawa et al. (Asogawa et al., 2020).

The frequently targeted STR loci were originally selected for analysis using CE. Hence, not all STR loci are equally suited for genotyping by sequencing. In order to assess why genotyping of some specific STR loci after nanopore sequencing is inaccurate, Tytgat et al. identified locus-dependent characteristics that compromise accurate genotyping (Tytgat et al., 2020). These findings can assist in constructing an STR panel suitable for nanopore sequencing. Firstly, the presence of homopolymers in the repeat region proved cumbersome. Besides, high repeat numbers, the presence of partial repeats, and complex repeat patterns showed to be problematic for nanopore sequencing. Lastly, sequence similarity between the repeat region and the flanking region hampers correct alignment to reference alleles. Nevertheless, genotyping a subset of STR loci after nanopore sequencing could reliably be performed for single-contributor samples. In the same report, Tytgat and colleagues compared the genotyping accuracy after sequencing using the R9.4 versus the R10 version of nanopores. Although sequencing using the R10 flow cell improved the genotyping accuracy for loci characterized by the presence of

homopolymers, the overall genotyping accuracy was lower due to an increased level of random sequencing errors.

More recent publications focus on using a commercially available NGS sequencing kit to perform amplification prior to nanopore sequencing. Ren et al. and Tytgat et al. used the ForenSeq DNA Signature Prep Kit (Ren et al., 2021; Tytgat et al., 2021), commercialized by Verogen. Hall and colleagues used the Promega PowerSeq 46GY System (Hall et al., 2021). Owing to the continuous improvement of nanopores, kits, and base-calling software, most STR loci can now accurately be genotyped after nanopore sequencing. Moreover, the added value of sequencing STR loci instead of length-based analysis is now harnessed. A successful demonstration of iso-allele genotyping was obtained after amplification using the ForenSeq DNA Signature Prep Kit (Tytgat et al., 2021). Hall and colleagues (2021) included the detection of SNPs within the flanking region in their data analysis pipeline, STRspy. More than 90% of the SNPs occurring in these flanking regions could be genotyped correctly in their study (Hall et al., 2021).

An additional important aspect of improving genotyping accuracy is the workflow used for data analysis. Most reports describe the construction of a reference database containing all possible alleles of all loci targeted by the assay. After base-calling, reads are assigned to their corresponding loci, either by primer recognition or mapping against the human reference genome. In the next step, the obtained reads are aligned against the database of reference alleles. Each read is assigned to the best-matching allele. Optionally, an alignment score filter can be employed to discard reads affected by amplification, sequencing, and alignment errors. The resulting read counts are then used for genotyping. Developing a robust, validated data-analysis workflow is crucial, as exemplified by Tytgat et al. (2021). Two workflows were compared on an identical dataset, resulting in pronounced differences in genotyping accuracy. A drawback to these workflows is their dependence on a reference database. All possible (iso-)alleles should be present in this database. The presence of an allele in the sample that is not recorded in the database will lead to incorrect genotyping. Expanding the existing databases containing STR sequences, such as STRSeq (Gettings et al., 2017), with sequence data gathered by the community will be crucial. Although a considerable improvement over the course of a few years can be noted, further advancements should be realized for this technology to be implemented in routine use. Moreover, large developmental validation studies are needed to establish locus-dependent data-analysis parameters, e.g., allelic imbalance cut-offs.

Single nucleotide polymorphism genotyping

SNPs are an interesting marker for forensic research, although they are less commonly investigated than STRs. SNPs are particularly useful for the analysis of highly degraded DNA, as they can be analyzed from short amplicons (Børsting et al., 2013). Moreover, they are less prone to mutation, and they can reveal some phenotypic traits of the sample

donor (Butler et al., 2007). Unfortunately, due to their biallelic nature, SNPs are characterized by a lower discriminatory power compared to STRs, therefore necessitating higher-order multiplexing (Butler et al., 2007). The SNPforID consortium developed a multiplex targeting 52 SNP markers, thereby approaching the discriminatory power of the routinely used STR panels consisting of 10–15 loci (Sanchez et al., 2006). Cornelis and colleagues used this panel to assess the accuracy of SNP genotyping based on nanopore reads (Cornelis et al., 2017). After sequencing using an R9.4 flow cell, all but one of the loci were genotyped correctly. However, for some loci, achieving robust genotyping after nanopore sequencing was problematic. All of these loci contained a homopolymer in the region flanking the targeted SNP, again indicating how troublesome homopolymers are for nanopore sequencing.

In a follow-up study, Cornelis and colleagues investigated a panel targeting 16 triallelic SNPs (Cornelis et al., 2019). Nonbinary SNP panels are characterized by a higher discriminatory power compared to biallelic SNP panels. Moreover, analyzing multiallelic SNPs enables the reliable detection of mixture samples (Phillips et al., 2004, pp. 27–29). Four out of five investigated samples could be genotyped correctly. Again, some problematic loci were identified, which should preferably be avoided during the design of panels dedicated to nanopore sequencing. This latest study by Cornelis and colleagues dates back to 2018. In the following years, the read accuracy of nanopore sequencing has improved. Two studies dating from 2021, as mentioned above, assessed the genotyping accuracy of nanopore sequencing after amplification using the ForenSeq DNA Signature Prep Kit (Ren et al., 2021; Tytgat et al., 2021). Besides autosomal and nonautosomal STRs, this kit targets 94 SNP loci, thereby combining the advantages of both types of markers. Both studies reported an SNP genotyping accuracy of $> 99\%$, indicating that nanopore sequencing is ideally suited to perform forensic SNP genotyping. Incorrect genotyping was mainly related to low coverage, which might be explained by the extremely high degree of multiplexing. Despite its higher error rate than Illumina sequencing, nanopore sequencing can accurately genotype forensic SNPs when sufficiently high coverage is obtained.

Mitochondrial DNA genotyping

As discussed above, human DNA profiling is mostly performed by analyzing STRs and SNPs present in nuclear DNA. However, in cases where nuclear DNA is highly degraded, analysis of mitochondrial DNA (mtDNA) might provide some valuable information. The mitochondrial genome is about 16.5 kb long and is characterized by a very high copy number per cell (Anderson et al., 1981). As mtDNA is maternally inherited, all maternal relatives share the same mtDNA sequence, apart from mutations. It should be noted that the mtDNA polymerase has lower fidelity, resulting in a higher mutation rate of mtDNA compared to nuclear DNA. The maternal inheritance makes mtDNA genotyping extremely useful for mass disaster victim identification, as maternal relatives

can donate a sample to serve as a reference, in case no reference DNA of the victim is available (Budowle et al., 2003). In most cases, the hypervariable regions I and II (HV1 and HV2) are assessed (Wilson et al., 1995). As nanopore sequencing is a long-read sequencing system, the entire mtDNA genome can be sequenced more easily, thereby reducing the error in haplogroup assignment (King et al., 2014; Parsons & Coble, 2001). This increased discriminatory power is particularly of interest for discerning a mixture sample from a heteroplasmic sample. Heteroplasmy is the condition in which a cell harbors a mixture of nonidentical mtDNA copies and is thus difficult to distinguish from a mixture sample (Stefano et al., 2017).

Only limited research has been conducted on mtDNA nanopore sequencing for forensic purposes. Lindberg and colleagues successfully phased variants of a 1:1 mixture sample by combining Illumina short-read sequencing and nanopore long-read sequencing (Lindberg et al., 2016). A few years later, when the nanopore sequencing technology was more mature, Zascavage and colleagues described nanopore sequencing of mtDNA without any preenrichment step, such as PCR (Zascavage et al., 2019). Omitting the PCR amplification step eliminates any sequence errors introduced during PCR and amplification bias. An overall error rate of approximately 1.00% was reported for the complete mitochondrial genome. A considerable portion of these errors could be attributed to homopolymeric stretches. From this, it is clear that nanopore sequencing is a promising tool for forensic mtDNA analysis.

Nonhuman forensics

Species identification

Species identification by DNA analysis is commonly used in a forensic setting, e.g., to aid in reconstructing a crime in which animals are involved or in cases of trafficking in protected species. To realize species identification, 'DNA barcoding' is mainly used. This methodology consists of a PCR step, during which standardized genetic regions are amplified, followed by sequencing of the resulting amplicons (Hebert et al., 2003). The obtained reads are then compared to a reference database containing the sequences of the standardized genetic regions of several species. This way, the sequences serve as a unique barcode for a specific species. Currently, sequencing is performed mainly by Sanger sequencing or NGS technologies. Nanopore sequencing offers some benefits to this field. First, long-read sequencing might increase the robustness of species identification workflows, which is needed to discriminate closely related species (Vasiljevic et al., 2021). Second, the portability of the MinION device enables species identification at remote locations, which can be useful in many cases. Lastly, the low cost of the platform enables laboratory-based sequencing in regions where access to costly NGS devices is limited.

As the error rate of nanopore reads is higher compared to most NGS methods, a well-performing data-analysis workflow is crucial to obtain highly accurate consensus

sequences, which can be compared to reference databases. Sahlin et al. developed a bioinformatics pipeline dedicated to species identification by third-generation sequencing strategies such as ONT and PacBio, called NGSpeciesID (Sahlin et al., 2021). Reads are first filtered based on quality and size, followed by a clustering step of similar reads. A consensus read is then constructed, followed by the merging of reverse-complement consensus reads. Finally, a polishing step is performed to obtain the final consensus sequence. Developmental validation of the NGSpeciesID pipeline after performing nanopore sequencing for forensic genetic species identification was performed by Vasiljevic and colleagues (Vasiljevic et al., 2021). The authors concluded that the construction of an accurate and reliable single consensus sequence is feasible using the MinION sequencer. Moreover, they consider the species identification results obtained using their workflow robust enough for forensic purposes.

Microbial analysis

Microbial forensics has been defined by Budowle et al. as "the scientific discipline dedicated to analyzing evidence from a bioterrorism act, biocrime, or inadvertent microorganism/toxin release for attribution purposes" (Budowle et al., 2003). Probably the most well-known example is the so-called anthrax letter attack. In 2001, letters containing spores of *Bacillus anthracis* (*B. anthracis*) were posted to, among others, two US senators. The attack resulted in 22 infections, of which 5 people died. Therefore, *B. anthracis* is a frequently targeted organism within the field of microbial forensics. Other high-priority pathogens are *Clostridium botulinum*, which causes botulism, and *Yersinia pestis (Y. pestis)*, which causes plague. All these pathogens have been analyzed by nanopore sequencing (Gonzalez-Escalona & Sharma, 2020). Gargis et al. performed whole-genome sequencing for *B. anthracis* and *Y. pestis* and could even detect different antimicrobial resistance genes (Gargis et al., 2019). One of the advantages of nanopore sequencing in this field is that real-time information can be obtained during sequencing. McLaughin and colleagues performed nanopore whole-genome sequencing of *B. anthracis* a couple of hours after receipt (McLaughlin et al., 2020). This clearly illustrates the potential value of nanopore sequencing in emergency situations where fast answers are needed. Although some of the abovementioned studies combine nanopore sequencing with a short-read sequencing platform to accurately analyze the genome, nanopore sequencing is capable of detecting the targeted pathogens.

Besides its applications in bioterrorism, microbial forensics is also applied to analyze the human microbiome in a forensic context. Ongoing research tries to retrieve a variety of forensic information from the microbiome of samples. The most used sample type for this application is touch samples, which are compared to the skin microbiome of an individual of interest (Schmedes et al., 2017). Second, postmortem interval estimation based on the microbiome has been explored, as repeatable shifts in microbial community composition occur during body decomposition (Metcalf, 2019). Next, fingerprinting the

microbiome detected in the soil might provide insights into geographical provenance (Giampaoli et al., 2014). In some instances, analysis of the microbiome can reveal the cause of death. An example is the detection of bacterioplankton as an indicator of drowning (Rutty et al., 2015). Lastly, the tissue or body fluid identity might be retrieved from the microbiome (Hanssen et al., 2017).

Several interesting potential forensic applications exist for microbiome sequencing. This field is challenging, because of the huge inter- and intraindividual variability (López et al., 2021). Nanopore sequencing could offer some technical advantages in this field as long reads deliver accurate identification of closely related species and can sequence informative genes (e.g., 16S rRNA) in their entirety, improving the resolution of identification (López et al., 2021). Unfortunately, to our knowledge, no research on the analysis of the human microbiome for forensic purposes has been performed yet by nanopore sequencing. Nevertheless, nanopore sequencing has proven to allow the analysis of complex microbial samples, as reviewed by Ciuffreda and colleagues (2021). Therefore, nanopore sequencing might play a role in the further development and validation of the abovementioned assays.

Opportunities

Messenger RNA analysis

The performance and possible advantages of nanopore sequencing over current techniques have not been researched in several niche forensic applications. A first example is messenger RNA analysis, which can be employed for body fluid identification. The composition of a biological sample found at a crime scene can reveal crucial information for an ongoing investigation. mRNA-based assays have been developed for the detection of semen, (menstrual) blood, cervicovaginal fluid, saliva, and skin, as described by Roeder et al. (Roeder & Haas, 2016, pp. 13–31). This is currently realized mainly by performing multiplex PCR, followed by profiling using CE. The nanopore platform offers cDNA kits to analyze mRNA found at a crime scene. A proof-of-concept of how nanopore sequencing can be used to perform body fluid identification was obtained by Dunn and Fleming (2019). They successfully identified salivary gland and tongue mRNAs by sequencing both direct mRNA and cDNA.

Epigenetics and epigenomics

An advantage of nanopore sequencing is its capability to directly distinguish methylated and unmethylated cytosines (Rand et al., 2017). DNA methylation analysis can be used to determine the cell or tissue type (Frumkin et al., 2011), but it can also be applied for age estimation (Park et al., 2016). Moreover, differentiation between monozygotic twins was obtained by analyzing the so-called epigenetic fingerprint (Lévesque et al., 2014). Interest in analyzing the epigenome for forensic purposes is growing. This should enable

lifestyle and environmental exposure prediction, thereby providing law enforcement services with extra-valuable information on unknown perpetrators.

Nanopore sequencing offers the advantage of analyzing nucleotide modifications without the need for chemical or enzymatic conversions, or PCR. Multiple tools have been developed to decipher the methylation state of a DNA or RNA molecule. A systematic benchmarking of these tools reveals pronounced differences in the prediction results of methylation frequencies between tools (Yuen et al., 2021). In this research, 10 regions of forensic interest were sequenced using a Cas9-targeted nanopore sequencing protocol. These regions included ancestry-informative SNPs, phenotypic-informative SNPs, and autosomal STRs. Although a lot of research is still required in both the accurate prediction of the methylation state and the interpretation of these data in a forensic context, nanopore sequencing is a promising tool for forensic epigenomic analysis.

Challenges

Nanopore sequencing is still a relatively new technology, as it was only introduced in 2014 (Deamer et al., 2016). It did not take long for the forensic field to start investigating the potential of this new technology. In 2016, Zaaijer and colleagues reported a fast method for the identification of DNA samples using the MinION sequencer by sequencing SNPs at low coverage without any prior amplification step. A Bayesian algorithm could then compare the obtained results to a reference database, taking into account the error rate of the sequencing platform (Zaaijer et al., 2016). A few years later, the focus of most research in the field shifted toward the identification of individuals by STR genotyping, which is still the most commonly used method for DNA profiling. The results obtained throughout the years by different research groups on STR genotyping by nanopore sequencing perfectly demonstrate the improvements realized by ONT. Whereas the first report could only obtain partial profiles due to the very high sequencing error rate (Cornelis et al., 2018), single-contributor samples can now successfully be analyzed (Hall et al., 2021; Tytgat et al., 2021). Nevertheless, some hurdles should be overcome for nanopore sequencing to find its way toward routine use in the field.

The first important challenge is the raw read accuracy, which is still lower than Illumina sequencing. This results in noisier genotyping data (Tytgat et al., 2020), which is undesirable in a forensic setting where often challenging samples, e.g., low-input or mixture samples, are analyzed. ONT is continuously increasing the raw read accuracy by improving its pores, sequencing kits, and base callers. At the time of writing, the R10.4 version was the most recent flow cell version, with a modal accuracy above 99% reported when used in combination with the so-called 'Q20+' sequencing kit. Moreover, a high consensus accuracy for homopolymeric regions was obtained (Sereika et al., 2022). Other interesting approaches to increase the read accuracy focus on the amplification step prior to sequencing. Rolling circle amplification results in multiple

copies of the target region within one read, allowing the construction of a consensus read based on these separate subreads originating from the same molecule in the original sample (Wilson et al., 2019). A second strategy is the use of unique molecular identifiers (UMIs), which allow for the creation of a consensus read based on all reads tagged with the same UMI (Karst et al., 2021). These results are promising for all possible applications of nanopore sequencing within the field of forensics.

A second challenge is the portability of the library preparation steps. Although the MinION device is small and perfectly portable, library preparation usually requires basic laboratory tools. In order to obtain a sample-in-answer-out system, ONT developed the VolTRAX, an automated sample and library preparation device. This device can mix reagents in a controlled way, perform bead purification, and perform quality control of the sample by measuring the UV absorption profile at 260 and 280 nm. According to ONT, future versions of the controlling software will enable PCR (VolTRAX V2 Technical Specification, 2021). Implementing the entire library preparation workflow, including PCR, will be crucial for on-site DNA profiling in a forensic setting.

Conclusions

Nanopore sequencing is a relatively new long-read sequencing technology. The MinION sequencer is a portable, low-cost device. These features offer many opportunities in the field of forensic DNA genotyping. Most research has focused on the identification of individuals by genotyping STRs and SNPs in view of the portability of the system, which would enable DNA fingerprinting at the crime scene. A second interesting application is whole mtDNA sequencing, exploiting the long nanopore reads that can span an entire mitochondrial genome. Third, nanopore sequencing can be used for DNA methylation detection without any prior enzymatic or chemical conversion step, allowing tissue type determination. Other interesting applications are nonhuman forensics and body fluid identification. The raw read accuracy has improved tremendously because of the continued improvement of the flow cells, the sequencing kits, and the base-calling software. Furthermore, the VolTRAX device will enable on-site library preparation with limited human intervention and virtually no need for other equipment. We believe that these innovations will push nanopore sequencing toward routine use in the field of forensic DNA genotyping.

References

Anderson, S., Bankier, A. T., Barrell, B. G., de Bruijn, M. H., Coulson, A. R., Drouin, J., Eperon, I. C., Nierlich, D. P., Roe, B. A., & Sanger, F. (1981). Sequence and organization of the human mitochondrial genome. *Nature, 290*(5806), 457–465.

Asogawa, M., Ohno, A., Nakagawa, S., Ochiai, E., Katahira, Y., Sudo, M., Osawa, M., Sugisawa, M., & Imanishi, T. (2020). Human short tandem repeat identification using a nanopore-based DNA sequencer: A pilot study. *Journal of Human Genetics, 65*(1), 21–24.

Børsting, C., Mogensen, H. S., & Morling, N. (2013). Forensic genetic SNP typing of low-template DNA and highly degraded DNA from crime case samples. *Forensic Science International: Genetics, 7*(3), 345–352.

Børsting, C., & Morling, N. (2015). Next generation sequencing and its applications in forensic genetics. *Forensic Science International: Genetics, 18*, 78–89.

Bruijns, B., Tiggelaar, R., & Gardeniers, H. (2018). Massively parallel sequencing techniques for forensics: A review. *Electrophoresis, 39*(21), 2642–2654.

Budowle, B., Allard, M. W., Wilson, M. R., & Chakraborty, R. (2003). Forensics and mitochondrial DNA: Applications, debates, and foundations. *Annual Review of Genomics and Human Genetics, 4*(1), 119–141.

Budowle, B., Schutzer, S. E., Einseln, A., Kelley, L. C., Walsh, A. C., Smith, J. A., Marrone, B. L., Robertson, J., & Campos, J. (2003). Building microbial forensics as a response to bioterrorism. *Science, 301*(5641), 1852–1853. American Association for the Advancement of Science.

Butler, J. M. (2005). *Forensic DNA typing: Biology, technology, and genetics of STR markers*. Elsevier.

Butler, J. M., Coble, M. D., & Vallone, P. M. (2007). Thoughts on the future of forensic DNA testing. *Forensic Science, Medicine and Pathology, 3*(3), 200–205.

Churchill, J. D., Schmedes, S. E., King, J. L., & Budowle, B. (2016). Evaluation of the Illumina® beta version ForenSeq™ DNA signature prep kit for use in genetic profiling. *Forensic Science International: Genetics, 20*, 20–29.

Ciuffreda, L., Rodríguez-Pérez, H., & Flores, C. (2021). Nanopore sequencing and its application to the study of microbial communities. *Computational and Structural Biotechnology Journal, 19*, 1497–1511.

Cornelis, S., Gansemans, Y., Deleye, L., Deforce, D., & Van Nieuwerburgh, F. (2017). Forensic SNP genotyping using nanopore MinION sequencing. *Scientific Reports, 7*(1), 1–5.

Cornelis, S., Gansemans, Y., Vander Plaetsen, A.-S., Weymaere, J., Willems, S., Deforce, D., & Van Nieuwerburgh, F. (2019). Forensic tri-allelic SNP genotyping using nanopore sequencing. *Forensic Science International: Genetics, 38*, 204–210.

Cornelis, S., Willems, S., Van Neste, C., Tytgat, O., Weymaere, J., Vander Plaetsen, A.-S., Deforce, D., & Van Nieuwerburgh, F. (2018). Forensic STR profiling using oxford nanopore technologies' MinION sequencer. *bioRxiv*, 433151.

Deamer, D., Akeson, M., & Branton, D. (2016). Three decades of nanopore sequencing. *Nature Biotechnology, 34*(5), 518–524.

Dunn, N., & Fleming, R. (2019). Investigating the potential of body fluid identification using the Oxford Nanopore MinION. *Australian Journal of Forensic Sciences, 51*(Suppl. 1), S27–S30.

Eskola, H., Väisänen, O., Viik, J., & Hyttinen, J. (2017). *EMBEC and NBC 2017: Joint conference of the European medical and biological engineering conference (EMBEC) and the nordic-baltic conference on biomedical engineering and medical physics (NBC), June 2017 (Vol. 65)*. Springer.

Frumkin, D., Wasserstrom, A., Budowle, B., & Davidson, A. (2011). DNA methylation-based forensic tissue identification. *Forensic Science International: Genetics, 5*(5), 517–524.

Gargis, A. S., Cherney, B., Conley, A. B., McLaughlin, H. P., & Sue, D. (2019). Rapid detection of genetic engineering, structural variation, and antimicrobial resistance markers in bacterial biothreat pathogens by nanopore sequencing. *Scientific Reports, 9*(1), 1–14.

Gettings, K. B., Borsuk, L. A., Ballard, D., Bodner, M., Budowle, B., Devesse, L., King, J., Parson, W., Phillips, C., & Vallone, P. M. (2017). STRSeq: A catalog of sequence diversity at human identification short tandem repeat loci. *Forensic Science International: Genetics, 31*, 111–117.

Giampaoli, S., Berti, A., Di Maggio, R., Pilli, E., Valentini, A., Valeriani, F., Gianfranceschi, G., Barni, F., Ripani, L., & Spica, V. R. (2014). The environmental biological signature: NGS profiling for forensic comparison of soils. *Forensic Science International, 240*, 41–47.

Gonzalez-Escalona, N., & Sharma, S. K. (2020). Closing *Clostridium botulinum* group I genomes using a combination of short-and long-reads. *Frontiers in Microbiology, 11*, 239.

Hall, C. L., Kesharwani, R. K., Phillips, N. R., Planz, J. V., Sedlazeck, F. J., & Zascavage, R. R. (2021). Accurate profiling of forensic autosomal STRs using the oxford nanopore technologies MinION device. *Forensic Science International: Genetics*, 102629.

Hanssen, E. N., Avershina, E., Rudi, K., Gill, P., & Snipen, L. (2017). Body fluid prediction from microbial patterns for forensic application. *Forensic Science International: Genetics, 30*, 10–17.

Hebert, P. D., Cywinska, A., Ball, S. L., & DeWaard, J. R. (2003). Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London. Series B: Biological Sciences, 270*(1512), 313–321.

Jain, M., Olsen, H. E., Paten, B., & Akeson, M. (2016). The oxford nanopore MinION: Delivery of nanopore sequencing to the genomics community. *Genome Biology, 17*(1), 1–11.

Karst, S. M., Ziels, R. M., Kirkegaard, R. H., Sørensen, E. A., McDonald, D., Zhu, Q., Knight, R., & Albertsen, M. (2021). High-accuracy long-read amplicon sequences using unique molecular identifiers with nanopore or PacBio sequencing. *Nature Methods, 18*(2), 165–169.

Kasianowicz, J. J., Brandin, E., Branton, D., & Deamer, D. W. (1996). Characterization of individual polynucleotide molecules using a membrane channel. *Proceedings of the National Academy of Sciences, 93*(24), 13770–13773.

King, J. L., LaRue, B. L., Novroski, N. M., Stoljarova, M., Seo, S. B., Zeng, X., Warshauer, D. H., Davis, C. P., Parson, W., & Sajantila, A. (2014). High-quality and high-throughput massively parallel sequencing of the human mitochondrial genome using the Illumina MiSeq. *Forensic Science International: Genetics, 12*, 128–135.

Lévesque, M. L., Casey, K. F., Szyf, M., Ismaylova, E., Ly, V., Verner, M.-P., Suderman, M., Brendgen, M., Vitaro, F., & Dionne, G. (2014). Genome-wide DNA methylation variability in adolescent monozygotic twins followed since birth. *Epigenetics, 9*(10), 1410–1422.

Lindberg, M. R., Schmedes, S. E., Hewitt, F. C., Haas, J. L., Ternus, K. L., Kadavy, D. R., & Budowle, B. (2016). A comparison and integration of MiSeq and MinION platforms for sequencing single source and mixed mitochondrial genomes. *PLoS One, 11*(12), e0167600.

López, C. D., Vidaki, A., & Kayser, M. (2021). Integrating the human microbiome in the forensic toolkit: Current bottlenecks and future solutions. *Forensic Science International: Genetics*, 102627.

McLaughlin, H. P., Bugrysheva, J. V., Conley, A. B., Gulvik, C. A., Cherney, B., Kolton, C. B., Marston, C. K., Saile, E., Swaney, E., & Lonsway, D. (2020). Rapid nanopore whole-genome sequencing for anthrax emergency preparedness. *Emerging Infectious Diseases, 26*(2), 358.

Metcalf, J. L. (2019). Estimating the postmortem interval using microbes: Knowledge gaps and a path to technology adoption. *Forensic Science International: Genetics, 38*, 211–218.

Park, J.-L., Kim, J. H., Seo, E., Bae, D. H., Kim, S.-Y., Lee, H.-C., Woo, K.-M., & Kim, Y. S. (2016). Identification and evaluation of age-correlated DNA methylation markers for forensic use. *Forensic Science International: Genetics, 23*, 64–70.

Parsons, T. J., & Coble, M. D. (2001). Increasing the forensic discrimination of mitochondrial DNA testing through analysis of the entire mitochondrial DNA genome. *Croatian Medical Journal, 42*(3), 304–309.

Phillips, C., Lareu, V., Salas, A., & Carracedo, A. (2004). Nonbinary single-nucleotide polymorphism markers. In *International congress series* (Vol. 1261, pp. 27–29). Elsevier.

Rand, A. C., Jain, M., Eizenga, J. M., Musselman-Brown, A., Olsen, H. E., Akeson, M., & Paten, B. (2017). Mapping DNA methylation with high-throughput nanopore sequencing. *Nature Methods, 14*(4), 411–413.

Rang, F. J., Kloosterman, W. P., & de Ridder, J. (2018). From squiggle to basepair: Computational approaches for improving nanopore sequencing read accuracy. *Genome Biology, 19*(1), 1–11.

Ren, Z.-L., Zhang, J.-R., Zhang, X.-M., Liu, X., Lin, Y.-F., Bai, H., Wang, M.-C., Cheng, F., Liu, J.-D., & Li, P. (2021). Forensic nanopore sequencing of STRs and SNPs using verogen's ForenSeq DNA signature prep kit and MinION. *International Journal of Legal Medicine*, 1–9.

Roeder, A. D., & Haas, C. (2016). Body fluid identification using mRNA profiling. In W. Goodwin (Ed.), *1420. Forensic DNA typing protocols* (pp. 13–31). New York, NY: Humana Press. https://doi.org/10.1007/978-1-4939-3597-0_2

Rutty, G. N., Bradley, C. J., Biggs, M. J., Hollingbury, F. E., Hamilton, S. J., Malcomson, R. D., & Holmes, C. W. (2015). Detection of bacterioplankton using PCR probes as a diagnostic indicator for drowning; the leicester experience. *Legal Medicine, 17*(5), 401–408.

Sahlin, K., Lim, M. C., & Prost, S. (2021). DNA barcode and amplicon consensus generation from long-read sequencing data. *Ecology and Evolution, 11*(3), 1392–1398.

Sanchez, J. J., Phillips, C., Børsting, C., Balogh, K., Bogus, M., Fondevila, M., Harrison, C. D., Musgrave-Brown, E., Salas, A., & Syndercombe-Court, D. (2006). A multiplex assay with 52 single nucleotide polymorphisms for human identification. *Electrophoresis, 27*(9), 1713–1724.

Schmedes, S. E., Woerner, A. E., & Budowle, B. (2017). Forensic human identification using skin microbiomes. *Applied and Environmental Microbiology, 83*(22), e01672–e01717.

Sereika, M., Kirkegaard, R. H., Karst, S. M., Michaelsen, T. Y., Sørensen, E. A., Wollenberg, R. D., & Albertsen, M. (2022). Oxford Nanopore R10. 4 long-read sequencing enables near-perfect bacterial genomes from pure cultures and metagenomes without short-read or reference polishing. *Nature Methods*, 823. https://doi.org/10.1038/s41592-022-01539-7

Stefano, G. B., Bjenning, C., Wang, F., Wang, N., & Kream, R. M. (2017). Mitochondrial heteroplasmy. *Mitochondrial Dynamics in Cardiovascular Medicine*, 577–594.

Tytgat, O., Gansemans, Y., Weymaere, J., Rubben, K., Deforce, D., & Van Nieuwerburgh, F. (2020). Nanopore sequencing of a forensic STR multiplex reveals loci suitable for single-contributor STR profiling. *Genes, 11*(4), 381.

Tytgat, O., Škevin, S., Deforce, D., & Van Nieuwerburgh, F. (2021). Nanopore sequencing of a forensic combined STR and SNP multiplex. *Forensic Science International: Genetics*, 102621.

Van der Verren, S. E., Van Gerven, N., Jonckheere, W., Hambley, R., Singh, P., Kilgour, J., Jordan, M., Wallace, E. J., Jayasinghe, L., & Remaut, H. (2020). A dual-constriction biological nanopore resolves homonucleotide sequences with high fidelity. *Nature Biotechnology, 38*(12), 1415–1420.

Vasiljevic, N., Lim, M., Humble, E., Seah, A., Kratzer, A., Morf, N. V., Prost, S., & Ogden, R. (2021). Developmental validation of oxford nanopore technology MinION sequence data and the NGSpeciesID bioinformatic pipeline for forensic genetic species identification. *Forensic Science International: Genetics, 53*, 102493.

VolTRAX V2 Technical Specification. (2021). v2-spec.pdf https://community.nanoporetech.com/requirements_documents/voltrax-v2-spec.pdf.

Wilson, M. R., DiZinno, J. A., Polanskey, D., Replogle, J., & Budowle, B. (1995). Validation of mitochondrial DNA sequencing for forensic casework analysis. *International Journal of Legal Medicine, 108*(2), 68–74.

Wilson, B. D., Eisenstein, M., & Soh, H. T. (2019). High-fidelity nanopore sequencing of ultra-short DNA targets. *Analytical Chemistry, 91*(10), 6783–6789.

Yuen, Z. W.-S., Srivastava, A., Daniel, R., McNevin, D., Jack, C., & Eyras, E. (2021). Systematic benchmarking of tools for CpG methylation detection from nanopore sequencing. *Nature Communications, 12*(1), 1–12.

Zaaijer, S., Gordon, A., Piccone, R., Speyer, D., & Erlich, Y. (2016). Democratizing DNA fingerprinting. *bioRxiv*, 061556.

Zascavage, R. R., Thorson, K., & Planz, J. V. (2019). Nanopore sequencing: An enrichment-free alternative to mitochondrial DNA sequencing. *Electrophoresis, 40*(2), 272–280.

Zhang, S., Bian, Y., Chen, A., Zheng, H., Gao, Y., Hou, Y., & Li, C. (2017). Developmental validation of a custom panel including 273 SNPs for forensic application using Ion Torrent PGM. *Forensic Science International: Genetics, 27*, 50–57.

CHAPTER 7

Microhaplotypes analysis for human identification using next-generation sequencing (NGS)

Mohammed H. Albujja
Department of Forensic Sciences, Naif Arab University for Security Sciences, Riyadh, Saudi Arabia

Introduction to forensic DNA analysis

Forensic DNA analysis is the process of analyzing and interpreting genetic materials to assist in investigating the identity of a perpetrator or victim, particularly in criminal investigations. Forensic DNA analysis involves the use of scientific techniques to extract, purify, and analyze DNA from biological samples collected from suspects, victims, and crime scenes. The analysis can take various forms, such as DNA sequencing, polymerase chain reaction (PCR), and STR (short tandem repeat) analysis. The sample types can include blood, tissue, bones, saliva, semen, hair, and other body fluids (Dash et al., 2020).

Once the DNA has been extracted and purified from the sample type, it is analyzed to produce a DNA profile of the individual whose DNA was collected. The DNA profile is generated by comparing the collected sample's genotype against a known reference sample, usually from a suspect or a database. The profiles can then be compared to identify relationships between people, such as siblings or parents and children, and to determine the presence or absence of a suspect at the crime scene (Butler, 2010c).

Forensic DNA analysis has played a vital role in solving crimes and implicating suspects. DNA evidence is considered the most reliable form of evidence because it is unique to each individual, and the technology can detect even minute quantities found at crime scenes. Furthermore, DNA evidence can provide conclusive evidence in instances where other forms of evidence may not be able to establish guilt beyond a reasonable doubt (Butler, 2010a).

The advent of DNA databases has further increased the significance of forensic DNA analysis in criminal investigations. Databases composed of DNA profiles from people convicted of certain crimes or submitted voluntarily have led to the identification of criminals who may otherwise have gone undetected. Additionally, the inclusion of DNA evidence has helped exonerate individuals who were wrongly accused or convicted of crimes (Butler, 2010b).

Next Generation Sequencing (NGS) Technology in DNA Analysis
ISBN 978-0-323-99144-5, https://doi.org/10.1016/B978-0-323-99144-5.00022-6
© 2024 Elsevier Inc.
All rights reserved.

DNA typing techniques used in forensic DNA analysis

Capillary electrophoresis (CE) is a powerful analytical technique that has revolutionized DNA analysis in forensic science. The introduction of CE in the early 1990s has significantly improved the speed, sensitivity, and accuracy of DNA analysis. CE can be used to separate and analyze complex mixtures of DNA samples in a single run, making it an efficient and cost-effective alternative to other traditional methods (Behl et al., 2021).

CE is based on the principle of electrokinetic separation, where charged molecules are separated by the application of an electric field across a narrow capillary tube filled with a buffer solution. The migration of molecules is determined by their charge-to-mass ratio, size, and shape. By adjusting the conditions of the electric field and buffer solution, different types of DNA fragments can be separated and analyzed.

CE is widely used in forensic DNA analysis for its ability to analyze small and degraded DNA samples, as well as mixed samples containing DNA from multiple sources. It is commonly used in the following areas:

1. STR analysis: Short Tandem Repeat (STR) analysis is a standard procedure for forensic DNA analysis. CE is used to separate STR alleles and obtain accurate genotypes.
2. Mitochondrial DNA analysis: CE can be used to analyze mtDNA fragments, which can be useful in cases where DNA is degraded or present in limited quantities.
3. Y-chromosome analysis: CE is used to analyze Y-chromosome-specific markers to determine the male ancestry of an individual.
4. Mixed DNA samples: CE can help to separate DNA from two or more individuals in a mixed sample to determine the contributor's profiles.

Advantages of capillary electrophoresis in forensic DNA analysis

The use of CE in forensic DNA analysis offers several advantages:

1. High resolution and sensitivity: CE can separate DNA fragments precisely, allowing for accurate identification of genotypes.
2. Cost-effectiveness: CE reduces the cost of analysis by enabling the separation and detection of multiple DNA fragments in a single run (multiplex amplification).
3. Efficient use of resources: With CE, researchers can analyze DNA samples with minimal volumes.

Fast turnaround time: CE can complete DNA analysis in a shorter time than traditional methods (Butler et al., 2004).

Limitations of capillary electrophoresis in forensic DNA analysis

Despite the numerous advantages, CE has some limitations in forensic DNA analysis, including:

1. Equipment costs: The cost of purchasing and maintaining the necessary equipment can be expensive.

2. Sample preparation: Proper sample preparation is critical since impurities can interfere with DNA separation and detection.
3. Inter-laboratory variation: There may be differences in CE protocols, reagents, and data analysis, which can lead to variations in results.

CE is a powerful tool for forensic DNA analysis and has become an essential part of DNA profiling in criminal investigations. With its high-resolution separation and sensitivity, CE offers significant advantages over traditional methods, especially in cases involving limited or mixed DNA samples. Nevertheless, optimization of protocols and continual quality assurance measures are necessary to achieve accurate and reproducible results (Fekete & Schmitt-Kopplin, 2007).

DNA sequencing

Next-Generation Sequencing (NGS) technologies have revolutionized the field of genomics by enabling the rapid, accurate, and cost-effective sequencing of whole genomes. This has led to a plethora of applications in various fields including personalized medicine, forensics, and population genetics (Hou et al., 2013). One such application is the analysis of microhaplotypes, which are small clusters of closely linked single nucleotide polymorphisms (SNPs) that can be used for various purposes including parentage testing, disease diagnosis, and population genetics studies. In contrast to STRs, microhaplotypes can be analyzed using NGS technology. This provides the ability to analyze multiple genetic markers at once, increasing the amount of information that can be gathered from a single DNA sample.

NGS-based microhaplotypes analysis is an emerging approach that harnesses the power of NGS to identify and analyze specific genetic markers known as microhaplotypes. Microhaplotypes are short DNA fragments that vary in length from 50 to 300 base pairs and contain several SNPs in close proximity. Due to their small size and high level of variability, microhaplotypes are ideal for use in forensic, evolutionary, and population genetics studies.

This technique involves using targeted sequencing to amplify and sequence specific microhaplotypes from DNA samples. This approach offers several advantages over traditional genotyping methods, including increased resolution, sensitivity, and accuracy, as shown in Table 7.1. By analyzing multiple microhaplotypes at once, researchers can obtain a more comprehensive view of an individual's genetic makeup and better understand the relationships between different populations.

As sequencing technologies continue to improve and new bioinformatic tools and standards are developed, we can expect to see even more exciting discoveries and applications in the years ahead. These advancements in molecular biology techniques have revolutionized the forensic science domain in the last few decades. Traditional methods such as polymerase chain reaction (PCR)-based SNP assays and STR typing have been widely used for forensic application. However, several limitations associated with these

Table 7.1 Comparison of NGS-based microhaplotypes analysis to traditional methods.

	NGS-based microhaplotypes analysis	PCR-based SNP assays	STR typing
Sample amount	Low (ng range)	High (μg range)	Low
Multiplexing	High (100−1000 markers in one run)	Low	Low
Resolution	High	High	Low
Sensitivity	High (detects minor contributors)	Low	Low
Discrimination	High	Low	High
Genomic region	Specific	Specific	Nonspecific

methods have led to the development of newer methods such as Next-Generation Sequencing (NGS)-based microhaplotypes analysis. In this chapter, I will compare and contrast the advantages and disadvantages of NGS-based microhaplotypes analysis with traditional DNA typing techniques.

Microhaplotypes as a tool for human identification

Definition of microhaplotypes

Human identification is an essential aspect of forensic science, and over the years, different molecular techniques have been employed to achieve this goal. Forensic DNA analysis plays a crucial role in criminal investigations and identifying missing persons; also, it is an essential tool for law enforcement agencies to identify suspects, link them to a crime scene, and exonerate innocent individuals. Short Tandem Repeats (STRs) have been the gold standard for human identification due to their multi-allelic nature and high variability (Butler, 2010d).

The development of NGS technologies has revolutionized the field of genomics and has opened up new avenues for forensic applications. One of the promising applications of NGS in forensic science is the identification of individuals based on their microhaplotypes (Wang et al., 2015).

However, implementing large-scale SNP assays in forensic labs still faces difficulties. Microhaplotypes (MH) or microhaps, consisting of nearby genetic markers, have emerged as a feasible and outstanding alternative to both STRs and Single Nucleotide Polymorphisms (SNPs) in human identification (Fan et al., 2022).

Microhaplotypes, as a novel tool for human identification, have been gaining increasing popularity in the field of forensic science (Turchi et al., 2017). With their high levels of variability and informativeness, microhaplotypes have been shown to be effective in resolving complex forensic cases and have the potential to improve the efficiency and accuracy of crime investigations. This chapter explores the various applications of microhaplotypes in forensic science, including those related to criminal

investigations, ancestry inference, and disaster victim identification. The potential benefits of microhaplotypes in forensic science make them a topic of great interest and importance for researchers, forensic experts, and law enforcement professionals alike (Staadig & Tillmar, 2021).

The technology has brought about a paradigm shift in genetic research, particularly in the field of forensic genetics, where it has enabled the development of new tools for human identification and ancestry inference. One such tool is microhaplotype analysis, which uses genetic markers that are shorter than traditional markers, making them more informative and discriminatory (Carratto et al., 2022; Turchi et al., 2017).

Microhaplotypes are short DNA fragments that are highly informative in identifying individuals and can be sequenced in a high-throughput manner using NGS. In the following paragraphs, we will discuss the utility of NGS for microhaplotypes analysis in human identification and explore the challenges and opportunities associated with this approach.

Types of compound genetic markers in microhaplotypes

Compound biomarkers consisting of two or more variants that occur within a small region can be regarded as generalized microhaplotypes, and include SNPs closely linked to STRs (SNP-STRs), insertion and deletion polymorphisms (indels) closely linked to STRs (DIP-STRs), indel polymorphisms closely linked to SNPs (DIP-SNPs). DIP-STRs and SNP-STRs are two types of compound genetic markers in microhaplotypes. DIP-STRs identified through MPS assays can be easily genotyped using CE platforms present in every forensic lab. Also, sequence-based information from SNP-STRs and DIP-STRs is important in designing allele-specific PCR primers, enabling the amplification of alleles from the minor DNA contributor that are absent in the genetic background of the major DNA source, thus eliminating the masking effect of the main DNA contributor (Jian et al., 2021).

Criminal investigations

One of the most important applications of microhaplotypes in forensic science is in criminal investigations. The high variability of microhaplotypes, coupled with their ability to provide reliable individual identification, make them an excellent tool for resolving complex cases. For instance, the analysis of microhaplotypes can allow forensic experts to differentiate between different individuals in cases where traditional DNA profiling fails due to sample degradation or complex mixtures of DNA (Fan et al., 2022). Furthermore, microhaplotypes can help resolve cases involving related individuals, such as siblings or half-siblings, where traditional DNA profiling may not provide sufficient discriminatory power. The allelic combinations of microhaplotypes can be observed in the same amplicon, even when only bi-allelic SNPs are present, providing multiple genotypes from a

single locus. This unique feature of microhaplotypes makes them highly efficient for the analysis of complex DNA mixtures.

Ancestry inference

Another area where microhaplotypes have shown great promise is in ancestry inference. By analyzing specific sets of microhaplotypes that are known to be associated with different ethnic groups, forensic experts can determine an individual's ancestral origin with a high degree of accuracy. This information can then be used to aid criminal investigations, as it can provide valuable clues about the possible identity of suspects.

Disaster victim identification

Microhaplotypes also hold great potential in disaster victim identification. In cases where traditional DNA profiling fails due to the degradation of DNA samples, microhaplotypes can be used to provide reliable and fast identification of victims. Furthermore, because microhaplotypes require relatively short DNA sequences, they can be amplified from degraded or damaged DNA samples, making them an ideal tool for identifying victims in mass disasters.

Challenges and limitations of microhaplotypes

Despite the advantages of NGS-based microhaplotypes, there are several challenges and limitations that need to be considered. One potential limitation is the need for large population databases containing microhaplotype data in order to establish reliable frequency estimates (Standage & Mitchell, 2020). Additionally, bioinformatics tools and algorithms for the analysis of microhaplotype data are still undergoing development, which may affect the accuracy and reliability of results.

One of the main challenges is the choice of the microhaplotype panel, which should be optimized for the specific research question and the target population. The panel should contain informative markers that allow for accurate haplotype phasing, while minimizing the genotyping error rate and the risk of missing variation (Kidd et al., 2018). Therefore, there is a need for standardization and validation of the methods and markers used for microhaplotypes analysis. The use of standardized protocols and markers is critical for the reproducibility and reliability of the results.

Another challenge is the bioinformatic processing, which can introduce biases and errors (Li et al., 2009). For example, alignment errors can lead to false-positive SNPs or haplotypes, while filtering criteria can exclude true variation or result in ascertainment bias. Therefore, it is important to apply rigorous quality control measures and to validate the results using independent methods. There is also a need for specialized bioinformatics tools and pipelines for the analysis of NGS data. The high volume of data generated by NGS can be overwhelming, and the development of efficient algorithms and analytical tools is essential for accurate and efficient analysis of microhaplotypes.

Finally, the interpretation of NGS-based microhaplotypes requires careful consideration of the genetic and evolutionary processes that shape the genetic variation. These processes can affect the pattern of linkage disequilibrium (LD), the distribution of allele frequencies, and the demographic history of the population. Therefore, it is crucial to integrate the genetic data with other sources of information, such as historical records, ecological data, and morphological traits (Bardan, 2019).

Although the use of microhaplotypes has been shown to be effective in many cases, there is still a need for further testing and validation before widespread adoption in forensic science.

Advantages of microhaplotypes over other DNA typing techniques

One of the main advantages of microhaplotypes is their high level of polymorphism, which allows for highly informative analysis of DNA samples. The random match probability (RMP) for a large panel of microhaplotypes can be as low as 10−100, making microhaplotypes a highly reliable method for individualization. Additionally, the heterozygosity of a microhaplotype can be measured as the effective number of alleles (Ae), with African populations having an Ae > 7.5 for some loci, and Native American populations having an Ae > 4.5 for a smaller panel of two dozen selected microhaplotypes (Kidd & Pakstis, 2022).

Microhaplotypes are also useful in ancestry inference, with at least 10 different ancestral clusters being definable by microhaplotypes when analyzed using structure. Furthermore, the Ae for a locus is identical to the Paternity Index (PI), which measures the informativeness of a locus in parentage testing. High Ae loci can also be useful in missing persons cases, where identifying a match with a close relative can help narrow down the search for the missing person (Kidd & Pakstis, 2022).

In mixture deconvolution, microhaplotypes offer a significant advantage over traditional STRs due to their high level of polymorphism. High Ae microhaplotypes can allow for the near certainty of seeing multiple additional alleles in a mixture of two or more individuals in a DNA sample. This makes microhaplotypes highly useful in cases where multiple contributors are present, such as in sexual assault cases (Kidd & Pakstis, 2022).

One of the main advantages of microhaplotypes in forensic DNA analysis is their ability to detect mixtures of multiple sources in a DNA sample. When all alleles are equally frequent, the probability of detecting three or more alleles in a mixture is at maximum. Effective number of alleles (Ae), a classical population genetics concept, can be used to convert unequal allele frequencies into a value equivalent to the number of equally frequent alleles, allowing microhaplotype loci to be ranked. The Ae value at a locus can convert unequal allele frequencies into an equivalent value of equally frequent alleles, allowing microhaplotype loci to be ranked. Microhaplotypes with Ae values of >3 are highly useful in ordinary forensic practice, with 3-SNP microhaplotypes

sometimes meeting this criterion, and 4-SNP microhaplotypes exceeding this criterion with values >4 (Tillmar & Phillips, 2017). Testing multiple microhaplotypes can exceed a 95% probability of detecting mixtures with as few as five loci with average Ae values slightly greater than 3.0 (Kidd & Speed, 2015).

Microhaplotypes can also be highly informative in forensic identification. Using a panel of higher Ae microhaplotypes, it is possible to outperform the standard CODIS markers.

Microhaplotypes for Biogeographical Ancestry: Using STRUCTURE, at least 10 different ancestral clusters can be defined by microhaplotypes, making them highly informative for biogeographical ancestry inference.

Applications of microhaplotypes in forensic DNA analysis

The massive parallel DNA sequencing (MPS) technologies are now making microhaplotypes a new type of forensic marker: a single sequence read can cover the expanse of the microhaplotypes and these loci become phase-known (Turchi et al., 2017, 2019).

Microhaplotypes are short DNA fragments that contain two or more closely linked SNPs within a small genomic region. These SNPs are highly polymorphic and can be used to differentiate individuals based on their unique haplotype profiles. Unlike other forensic markers such as STRs and SNPs, microhaplotypes are not affected by mutation, genetic drift, or recombination, and hence, offer a stable and reliable marker for human identification. They are useful for noninvasive prenatal paternity tests using cell-free fetal DNA (cffDNA) (Carratto et al., 2022).

Next-generation sequencing for microhaplotypes analysis

NGS has revolutionized the field of genomics and the way we understand genetic diversity and evolution. Enabling high-throughput sequencing of whole genomes or targeted regions, providing a wealth of information about population dynamics, genetic diversity, and phylogenetic relationships.

One of the most promising applications of NGS is the analysis of microhaplotypes. The use of NGS for microhaplotypes analysis offers several advantages over traditional methods. First, NGS allows for the simultaneous sequencing of thousands of microhaplotypes, providing a more comprehensive and informative profile for human identification. Second, NGS can provide higher accuracy and precision in the sequencing of microhaplotypes, reducing the possibility of errors and false identifications. Third, NGS can also identify novel and rare microhaplotypes, which may not be detected using traditional methods (Bruijns et al., 2018).

Microhaplotypes are short DNA fragments that are typically less than 300 base pairs (bp) in length and contain multiple SNP markers (Kidd et al., 2022; Kidd & Pakstis, 2022; Staadig & Tillmar, 2021). These SNPs are situated along the DNA molecule in a manner

that allows them to be inherited together as a block, leading to high levels of LD. This characteristic makes microhaplotypes ideal for forensic applications, as they can provide greater discrimination than traditional markers in certain populations.

Due to the small size of microhaplotypes, they are amenable to low-coverage sequencing, making them cost-effective and efficient for genetic analysis. Microhaplotypes have several advantages over individual SNPs, including higher resolution for distinguishing closely related individuals, greater power to detect recombination events, and lower genotyping error rates. In recent years, there has been growing interest in using microhaplotypes for forensic applications, population genetics, and medical research (Turchi et al., 2017).

Workflow of NGS-based microhaplotypes analysis

The workflow for NGS-based microhaplotypes analysis involves several key steps, including sample preparation, library preparation, sequencing, alignment, and data analysis. Each step has its considerations and challenges, which require careful attention to ensure optimal results. By following best practices and using appropriate methods and tools, researchers can obtain high-quality sequencing data and derive meaningful insights from their analyses. In this section, we will provide an overview of each of these steps and discuss some of the challenges and considerations involved in each step (Gandotra et al., 2020).

Sample preparation

The first step in the workflow is sample preparation, which involves collecting a biological sample from the individual of interest and extracting DNA. There are various methods available for DNA extraction, including manual and automated protocols. The choice of extraction method depends on the quality and quantity of the starting material, the downstream applications, and the laboratory resources.

One of the critical considerations in sample preparation is minimizing contamination and preserving the integrity of the DNA sample. Contamination can lead to false-positive results, while degraded DNA can affect the quality of the sequencing data. To ensure optimal sample quality, it is essential to follow best practices, such as using sterile equipment, minimizing handling, and storing samples at the appropriate temperature and humidity.

Library preparation for NGS-based microhaplotypes analysis

Once the DNA has been extracted, the next step is library preparation, which involves fragmenting the DNA and attaching adapters that enable sequencing on the NGS platform. Several methods are available for DNA fragmentation, including sonication, enzymatic digestion, and mechanical shearing. The choice of fragmentation method depends on the size and quality of the DNA sample and the downstream application.

After fragmentation, the DNA fragments are ligated with adapters that contain specific barcodes that enable multiplexing of multiple samples in a single sequencing run. Multiplexing can increase the throughput and efficiency of the sequencing process by reducing the number of sequencing runs needed to analyze a large set of samples (Gandotra et al., 2020).

Sequencing

After library preparation, the next step is sequencing, which involves loading the libraries onto the NGS platform and acquiring sequence data. The choice of sequencing platform and sequencing depth depends on the research question, the type of analysis, and the available resources. For microhaplotype analysis, low-coverage sequencing is typically sufficient, which reduces the cost and time required for sequencing (Wang et al., 2015).

Alignment

After sequencing, the next step is alignment, which involves mapping the sequence reads back to a reference genome or a custom reference panel. Alignment is necessary to determine the location and identity of the SNP markers within the microhaplotypes. Several software programs are available for alignment, such as BWA, Bowtie, and HISAT2. The choice of alignment software depends on the type of analysis and the sequencing platform used (Li et al., 2009).

Data analysis for NGS-based microhaplotypes analysis

The final step in the workflow is data analysis, which involves calling and interpreting the SNP markers within the microhaplotypes. The analysis can be performed using various software programs, such as GATK, FreeBayes, and SAMtools. The analysis typically involves filtering the SNPs based on quality scores, allele frequency, and linkage disequilibrium. The filtered SNPs are then combined into haplotypes, which can be used for downstream applications, such as forensics or population genetics (Pang et al., 2020).

Validation of NGS-based microhaplotypes analysis

Data analyzing of microhaplotypes requires careful consideration of the data quality, sequencing depth, and bioinformatic processing steps. In this section, I will describe the workflow for Data analysis of NGS-based microhaplotypes, including quality control, read alignment, SNP calling, haplotype reconstruction, and statistical inference. I will also discuss the challenges and limitations of NGS-based microhaplotypes and strategies for validating the results (Pang et al., 2020).

Quality control

The first step in analyzing NGS-based microhaplotypes is to assess the quality of the raw sequencing data. This includes evaluating the read length, base quality, and overall

sequencing depth. Low-quality reads should be removed, and adapter sequences should be trimmed to ensure accurate read alignment. In addition, technical replicates and positive controls should be included to monitor the reproducibility and accuracy of the sequencing results.

Read alignment

The next step is to align the cleaned reads to a reference genome or a custom-designed panel of microhaplotype sequences. Several popular alignment tools, such as BWA, Bowtie, and STAR, are available for this purpose. The choice of the alignment tool depends on the specific research question and the characteristics of the sequencing data, such as read length, sequencing depth, and the presence of indels.

SNP calling

Once the reads are aligned, the next step is to identify SNPs within the microhaplotypes. This can be done using various SNP callers, such as GATK, FreeBayes, and SAMtools. The quality of the SNP calls should be evaluated based on several criteria, such as the read depth, mapping quality, and allele frequency. False-positive SNPs can arise from sequencing errors, PCR duplicates, or rare variants. Therefore, it is important to filter out low-quality SNPs or those with low coverage.

Haplotype reconstruction

After SNP calling, the microhaplotypes can be phased and reconstructed into haplotypes using several methods, such as BEAGLE, HaploView, and SHAPEIT. Phasing refers to the process of determining which alleles belong to the same chromosome, while haplotype reconstruction infers the sequence of the two haplotypes for each individual. The accuracy of haplotype reconstruction depends on the number of SNPs per microhaplotype, the sequencing depth, and the LD structure of the markers. LD refers to the nonrandom association between nearby markers and can affect the accuracy of haplotype phasing.

Statistical inference

Once the haplotypes are reconstructed, they can be used for various downstream analyses, such as population structure, admixture, and relatedness estimation. These analyses rely on statistical models that account for the uncertainty in haplotype phasing and genotype calling. Several software packages, such as STRUCTURE, ADMIXTURE, and KING, are available for these analyses. It is important to validate the results using independent datasets or alternative methods, such as SNP genotyping, microsatellite typing, or mitochondrial DNA sequencing.

Challenges and limitations related to data analysis

Despite their promise, there are still challenges associated with microhaplotype analysis. For one, there is no consensus on how to define and select microhaplotypes (Kidd et al.,

2018). Different studies use different methods and criteria for defining microhaplotypes, which can lead to inconsistencies in results. Moreover, microhaplotype analysis requires complex statistical analyses and interpretation, which can be difficult for nonexperts.

As with any new technology, there are limitations and challenges that need to be addressed in order to improve the quality and reliability of the data generated by NGS-based microhaplotypes analysis.

NGS-based microhaplotype analysis still faces several challenges and limitations. One major challenge is the need for accurate and efficient bioinformatic pipelines to process and analyze the large amounts of sequencing data generated by NGS platforms. These pipelines must be able to handle errors and biases introduced by PCR amplification, sequencing chemistry, and other sources, while also detecting and filtering out artifacts such as chimeras and contaminants. Developing robust bioinformatic pipelines is an ongoing area of research, and several approaches have been proposed, including alignment-based methods, de novo assembly, and machine learning algorithms.

Another challenge for NGS-based microhaplotype analysis is the need to validate and standardize the assays used for SNP selection and PCR amplification. There is currently no standardized method for selecting and validating microhaplotype assays, and different approaches may produce different results depending on factors such as sample size, SNP density, and genotyping platforms. To address this issue, several consortia and working groups have been formed to develop guidelines and best practices for microhaplotype analysis, including the MicroHapDB consortium and the Forensic Genomics Consortium (Standage & Mitchell, 2020).

Future directions and opportunities in NGS-based microhaplotypes analysis

Despite these challenges, the use of NGS for microhaplotypes analysis in human identification presents several opportunities. The ability to sequence thousands of microhaplotypes simultaneously using NGS can provide a more comprehensive and informative profile for human identification, which can improve the accuracy and reliability of the results. NGS can also identify novel and rare microhaplotypes, which can increase the power of discrimination and improve the resolution of the analysis. Additionally, NGS can provide a more efficient and cost-effective method for the analysis of microhaplotypes, which can be particularly useful in large-scale forensic investigations.

The promises of microhaplotypes and advantages over traditional markers

Microhaplotypes offer several advantages over traditional genetic markers, as shown in Table 7.1. First, they are shorter than other markers, which makes them less prone to degradation and more stable in degraded DNA samples.

Second, they avoid the problems commonly seen in STR assays, such as stutter peaks/reads, which sometimes complicate Massive Parallel Sequencing (MPS) technology in forensic genetics. Stutter peaks/reads can lead to difficulties in the identification and deconvolution of mixtures when the difference between the major and minor DNA donors is high. This phenomenon occurs when the polymerase enzyme responsible for replicating the DNA strand during amplification pauses, creating a shorter fragment with one fewer repeat unit. This can lead to difficulties in accurately identifying alleles and may compromise the reliability of DNA analysis. On the other side, microhaplotypes do not exhibit stutter peaks/reads, making them easier to analyze and interpret in forensic DNA analysis (Oldoni et al., 2020; Standage & Mitchell, 2020).

Third, they can provide higher levels of discrimination than traditional markers, particularly in genetically diverse populations. This is because microhaplotypes consist of nearby genetic markers that can provide multiple genotypes from a single locus, even when only biallelic SNPs are present. This provides more information from a single DNA sample than SNPs can offer (Carratto et al., 2022).

Fourth, they are more informative, as they can provide information about both the maternal and paternal sides of an individual's family tree.

Fifth, they provide higher resolution, increased accuracy, and greater throughput. Analysis enables researchers to observe genetic variation at an unprecedented level of detail, providing insights into population genetics, phylogenetics, and forensic investigations.

Sixth, microhaplotypes can also be used to identify individuals, genetic disorders, and disease susceptibility factors, which could lead to more effective medical treatments and personalized therapies. Sixth, they are highly sensitive, requiring as little as 50 pg of DNA sample which is equivalent to the DNA found in eight cells; this makes them ideal for analyzing degraded DNA samples from challenging forensic contexts (Oldoni et al., 2020). In addition, it can differentiate mixtures with 2 contributors up to a 49:1 ratio and up to 6 DNA contributors; additionally, the number of contributors can be accurately defined in 2- to 4-person mixtures in more than 95% of the evaluated samples, and more than 83% accuracy for 2- to 6-person mixtures (Yang et al., 2022). This is an important feature in forensic investigations where DNA samples are often mixtures of genetic material from multiple individuals. These properties make microhaplotypes an outstanding alternative to both SNPs and STRs in human identification. To deal also with nonhuman DNA analysis and biogeographical ancestry cases (Pang et al., 2020).

Eighth, Microhaplotypes have the potential to identify biogeographical ancestry. A good set of Ancestry-Informative Markers (AIMs) organized into microhaplotypes would grant not only ancestry information but also human identifying potential as well. The major worldwide population groups (Africans, South-West Asians and Europeans, East-Asians, Americans, and the Pacific Islanders) can be defined with only 31 microhaplotypes (Oldoni et al., 2020).

This means that microhaplotypes could be used to provide both ancestry information and human identifying potential, which is particularly relevant in forensic investigations (Carratto et al., 2022).

Challenges of NGS-based microhaplotypes technique

Low coverage depth and variability

One of the main challenges in NGS-based microhaplotypes analysis is the low coverage depth and variability of the data generated by the technology. This can result in missing data, inaccurate genotyping, and false positives or negatives. The variability of the data also makes it difficult to compare results across different samples or platforms.

Read errors and artifacts

Another challenge is the presence of read errors and artifacts that can result in incorrect genotyping or false positives. These errors can result from sequencing errors, PCR biases, or sample contamination.

Data management and analysis

Managing and analyzing the large amounts of data generated by NGS-based microhaplotypes analysis can also be challenging. This requires sophisticated computational tools and expertize in bioinformatics and statistical analysis.

Improvements to NGS-based microhaplotypes analysis

Enhancing coverage depth and consistency

To improve the accuracy and reliability of NGS-based microhaplotypes analysis, one approach is to increase the coverage depth and consistency of the data generated by the technology. This can be achieved through various methods including sample pooling, targeted capture, and hybridization-based sequencing.

Reducing read errors and artifacts

To reduce the impact of read errors and artifacts on the accuracy of NGS-based microhaplotypes analysis, various approaches can be used including error correction algorithms, quality control measures, and validation of results using independent methods.

Developing and optimizing computational tools

To manage and analyze the large amounts of data generated by NGS-based microhaplotypes analysis, advanced computational tools need to be developed and optimized. This requires expertize in bioinformatics, statistics, and computer science.

There are challenges and limitations that need to be addressed to improve the accuracy and reliability of the data generated by the technology. By enhancing the coverage depth and consistency, reducing read errors and artifacts, and developing advanced computational tools, we can improve the quality and efficiency of

NGS-based microhaplotypes analysis and enable its wider adoption in research and clinical settings.

Future directions in microhaplotype analysis

Despite the challenges, microhaplotype analysis holds great promise for the future of forensic genetics (Carratto et al., 2022; Turchi et al., 2017). Here are some future directions that researchers should explore:

1. Defining microhaplotypes: Since there is currently no consensus on how to define microhaplotypes, researchers should work toward establishing a standard definition and selection criteria for microhaplotypes (Kidd & Speed, 2015).
2. Integrating microhaplotypes with other tools: Microhaplotype analysis could be combined with other next-generation sequencing—based tools, such as ancestry inference and phenotype prediction, to obtain even more powerful and informative results (Kidd et al., 2022).
3. Automation and standardization of analysis: With advances in automation technology, it should be possible to automate most of the data analysis and interpretation steps of microhaplotype analysis. Additionally, standardization of protocols and analytical pipelines could help to improve the reproducibility and comparability of results across different laboratories.

While the field still faces challenges such as defining microhaplotypes and interpreting results, advances in technology and collaboration can help drive the field forward into a future where microhaplotype analysis becomes a staple of forensic genetics research.

Ethical considerations for NGS-based microhaplotypes analysis

As with any new technology, NGS-based microhaplotypes analysis raises ethical considerations that must be taken into account to ensure the responsible and fair use of this powerful approach. In this section, we will explore the ethical issues surrounding NGS-based microhaplotypes analysis. We will examine the potential risks and benefits of this technology, discuss the obligations of scientists, researchers, and policymakers, and propose guidelines for the responsible use of NGS-based microhaplotypes analysis (Martinez-Martin & Magnus, 2019).

Risks of NGS-based microhaplotypes analysis

As with any technology, there are risks associated with this application, such as privacy concerns, discrimination, and the potential for misuse. The information obtained from NGS-based microhaplotypes analysis could be used to discriminate against individuals on the basis of their genetic profile, leading to potential harms such as loss of employment, denial of insurance coverage, or social ostracism. In the wrong hands, genetic data could also be used for nefarious purposes, such as identity theft, fraudulent activities, or targeted marketing (Butler, 2023).

Obligations of scientists, researchers, and policymakers

Scientists, researchers, and policymakers have a moral obligation to ensure that NGS-based microhaplotypes analysis is used in a responsible and ethical manner. This includes respecting the privacy and autonomy of individuals, protecting sensitive genetic information from unauthorized access, and promoting equitable and fair access to genetic services. To achieve these goals, it is essential to establish clear guidelines and standards for the collection, storage, and use of genetic data, as well as to educate individuals about the risks and benefits of genetic testing (Davey, 2014).

Guidelines for the responsible use of NGS-based microhaplotypes analysis

To ensure the responsible and ethical use of NGS-based microhaplotypes analysis, we propose the following guidelines:

1. Informed consent: Individuals must provide informed consent before any genetic testing is conducted, and they should be provided with clear and comprehensive information about the risks and benefits of the testing, as well as their rights to privacy and confidentiality.
2. Privacy and confidentiality: Genetic data must be kept confidential and protected from unauthorized access. Appropriate security measures must be implemented to safeguard sensitive information.
3. Data sharing: Genetic data should be shared only for legitimate research purposes, and proper protocols and safeguards should be implemented to minimize the risk of misuse.
4. Nondiscrimination: Genetic information must not be used to discriminate against individuals on the basis of their genetic profile, and appropriate legal protections should be put in place to prevent such discrimination.
5. Public education: The public must be educated about the risks and benefits of genetic testing, including the importance of protecting privacy and confidentiality.

To ensure that this technology is used in a responsible and ethical manner, it is essential to address the potential risks and benefits, and to establish clear guidelines and standards for the collection, storage, and use of genetic data. With careful consideration and responsible practices, NGS-based microhaplotypes analysis can contribute to scientific progress while respecting the privacy and autonomy of individuals (National DNA Database Ethics Group, 2017).

Potential applications of NGS-based microhaplotypes analysis

One of the key applications of NGS-based microhaplotypes analysis is in forensic science. Microhaplotypes have been shown to be highly informative for individual identification, making them valuable tools for criminal investigations and paternity testing. In addition,

microhaplotypes can be used to determine the ancestry of an individual, which can provide important contextual information for forensic investigations.

NGS-based microhaplotypes analysis also has important implications for evolutionary and population genetics research. By analyzing microhaplotypes from multiple individuals or populations, researchers can gain insights into patterns of genetic diversity and identify regions of the genome that are under selective pressure. This information can be used to better understand the genetic basis of complex traits and diseases and to develop targeted approaches for disease diagnosis and treatment.

Microhaplotypes have emerged as a promising tool for identifying biogeographical ancestry. Ancestry informative markers (AIMs) are genetic markers that vary in frequency between different populations and can be used to infer an individual's ancestry. Traditionally, AIMs have been identified using SNPs or STRs, but microhaplotypes offer a new alternative for the identification of ancestry. A good set of AIMs organized into microhaplotypes would grant not only ancestry information but also human identifying potential.

Several studies have shown that microhaplotypes can accurately identify major population groups. In a study by Oldoni et al. (2019), the authors identified 31 microhaplotypes that could define the major worldwide population groups, including Africans, South-West Asians and Europeans, East-Asians, Americans, and Pacific Islanders. These microhaplotypes were able to correctly assign individuals to their respective population group with an accuracy of over 99%.

The ability of microhaplotypes to identify biogeographical ancestry can have important applications in forensic investigations. For example, in cases where the identity of an individual is unknown, analysis of their DNA could provide information about their likely ancestry. This information could be used to narrow down the pool of potential suspects and help focus the investigation.

Microhaplotypes can also be useful in cases where human remains are found and there are no immediate clues to the identity of the individual. Analysis of the DNA from the remains could provide information about their ancestry, which could be used to help identify the individual. Additionally, the use of microhaplotypes could be particularly useful in cases where the individual's ancestry is not immediately clear, such as cases involving individuals with mixed ancestry.

Furthermore, microhaplotypes can also be used in combination with other forensic DNA analysis techniques, such as STRs, to increase the power of human identification. In a study by Zhu et al. (2019), the authors demonstrated that a combination of 28 microhaplotypes and 20 STRs provided high discriminatory power and ancestry information for a Chinese population sample. This combined approach could provide greater accuracy in identifying individuals and could be particularly useful in cases where traditional methods of identification have failed.

Future directions of NGS-based microhaplotypes analysis

In this section, we will explore the current state-of-the-art in NGS-based microhaplotype analysis, highlight key challenges and opportunities, and outline future directions for research in this exciting field.

NGS-based microhaplotype analysis: state-of-the-art

The first step in NGS-based microhaplotype analysis is the selection of informative SNPs that are suitable for multiplex PCR amplification and subsequent sequencing. Several methods have been developed for SNP selection, ranging from genome-wide association studies (GWAS) to targeted panels of SNPs based on functional annotation or population-specific frequencies. The most efficient approach for SNP selection is likely to involve a combination of these methods, taking into account factors such as linkage disequilibrium, allele frequency, and quality control metrics. Once informative SNPs have been identified, they can be amplified by multiplex PCR, and the resulting amplicons can be sequenced using NGS platforms such as Illumina or PacBio (Kidd & Pakstis, 2022).

NGS-based microhaplotype analysis offers several advantages over traditional methods such as capillary electrophoresis and Sanger sequencing. First, NGS enables the analysis of thousands or even millions of individual readouts per sample, providing much higher resolution than previous methods. Second, NGS allows for the detection of rare variants and low-frequency alleles that may be missed by other techniques. Third, NGS-based microhaplotype analysis requires relatively small amounts of DNA, making it feasible to use degraded or low-quality samples such as those obtained from forensic investigations or archeological remains.

Opportunities for future research in NGS-based microhaplotype analysis

In a Staadig and Tillmar study, the authors evaluated the performance of a custom-made QIAseq Microhaplotype panel that includes 45 different microhaplotype loci. The panel was evaluated using 75 samples of Swedish origin, and the haplotype frequencies were established. The authors also conducted sensitivity studies to determine the limit of detection of haplotypes and found that they could detect haplotypes at input amounts down to 0.8 ng. Mixture samples were also tested, and haplotypes for the minor contributor were detectable down to the level of 1:100. This demonstrates the high sensitivity of microhaplotypes and their potential use in mixture analysis, which is a common problem in forensic practice. Furthermore, the authors performed kinship simulations to evaluate the usefulness of the panel in kinship analysis. The results showed that both paternity and full sibling cases can clearly be solved. The authors noted that there were some primer design issues that could be optimized to increase the power of the assay. Nonetheless, the results of this study are promising for further implementation of this microhaplotype assay into the forensic field (Staadig & Tillmar, 2021).

Finally, there are several opportunities for future research in NGS-based microhaplotype analysis. One promising area involves the use of microhaplotypes for population genetics and evolutionary studies. By analyzing microhaplotype variation in large and diverse populations, researchers can gain insights into the evolutionary history and migration patterns of human and animal species, as well as identify regions of the genome under selection. Another area of research involves the development of new sequencing technologies and platforms that can provide even higher resolution and accuracy than current methods. For example, nanopore sequencing and single-molecule sequencing technologies offer the potential for real-time, long-read sequencing of individual molecules, enabling the detection of structural variants, haplotype phasing, and epigenetic modifications.

Conclusion

In conclusion, next-generation sequencing has revolutionized the field of genomics by providing researchers with the ability to sequence large amounts of genetic material quickly and accurately.

Microhaplotypes have emerged as a feasible alternative to both STRs and SNPs in forensic DNA analysis. They offer advantages in sensitivity, exclusion power, and discrimination power. A thorough selection of microhaplotypes can establish a single MPS assay to deal with a variety of forensic DNA analysis needs. Additionally, compound genetic markers such as DIP-STRs and SNP-STRs are getting attention in microhaplotypes analysis, making them a promising tool for forensic DNA analysis. Their high level of polymorphism allows for highly informative analysis of DNA samples, with a low RMP and high Ae values in certain populations.

As technology advances and more research is conducted in this field, it is likely that microhaplotypes will become even more important in forensic DNA analysis.

Microhaplotypes have shown great potential in identifying biogeographical ancestry. They offer a new alternative to traditional methods of identifying AIMs and can accurately define major worldwide population groups with a small set of markers. Microhaplotypes have important applications in forensic investigations and can be used in combination with other DNA analysis techniques to increase the accuracy of human identification.

However, there are still some limitations and challenges to the use of microhaplotypes in forensic DNA analysis. Further studies are needed to explore the potential of microhaplotypes in various aspects of forensic DNA analysis, such as mixture interpretation, quality control, and validation of new assays. Additionally, there are issues regarding data interpretation, privacy concerns, and ethical considerations that should be addressed.

References

Bardan, F. (2019). *New forensic DNA profiling techniques for human identification [A thesis, University of Adelaide].* https://digital.library.adelaide.edu.au/dspace/bitstream/2440/120161/1/Bardan2019_PhD.pdf.

Behl, A., Mishra, A., & Mishra, I. K. (2021). Tools and techniques used in forensic DNA typing. In *Handbook of DNA profiling* (pp. 1–31). Springer Singapore. https://doi.org/10.1007/978-981-15-9364-2_4-1

Bruijns, B., Tiggelaar, R., & Gardeniers, H. (2018). Massively parallel sequencing techniques for forensics: A review. In *Electrophoresis* (pp. 2642–2654). Wiley-VCH Verlag. https://doi.org/10.1002/elps.201800082

Butler, J. M. (2010a). Basics of DNA biology and genetics. In *Fundamentals of forensic DNA typing* (pp. 19–41). Elsevier. https://doi.org/10.1016/B978-0-12-374999-4.00002-3

Butler, J. M. (2010b). DNA databases. In *Fundamentals of forensic DNA typing* (pp. 259–289). Elsevier. https://doi.org/10.1016/B978-0-12-374999-4.00012-6

Butler, J. M. (2010c). DNA extraction. In *Fundamentals of forensic DNA typing* (pp. 99–109). Elsevier. https://doi.org/10.1016/B978-0-12-374999-4.00005-9

Butler, J. M. (2010d). Short tandem repeat markers. In *Fundamentals of forensic DNA typing* (pp. 147–173). Elsevier. https://doi.org/10.1016/B978-0-12-374999-4.00008-4

Butler, J. M. (2023). Recent advances in forensic biology and forensic DNA typing: INTERPOL review 2019–2022. *Forensic Science International: Synergy, 6*. https://doi.org/10.1016/j.fsisyn.2022.100311

Butler, J. M., Buel, E., Crivellente, F., & Mccord, B. R. (2004). Forensic DNA typing by capillary electrophoresis using the ABI Prism 310 and 3100 genetic analyzers for STR analysis. *Electrophoresis, 25*, 1397–1412. https://doi.org/10.1002/elps.200305822

Carratto, T. M. T., Moraes, V. M. S., Recalde, T. S. F., de Oliveira, M. L. G., & Mendes-Junior, C. T. (2022). Applications of massively parallel sequencing in forensic genetics. *Genetics and Molecular Biology, 45*(3). https://doi.org/10.1590/1678-4685-GMB-2022-0077

Dash, H. R., Shrivastava, P., & Das, S. (2020). *Introduction to forensic DNA analysis* (pp. 3–11). https://doi.org/10.1007/978-1-0716-0274-4_1

Davey, S. (2014). Next generation sequencing: Considering the ethics. *International Journal of Immunogenetics* (Vol 41,(6), 457–462. https://doi.org/10.1111/iji.12155

Fan, H., Xie, Q., Wang, L., Ru, K., Tan, X., Ding, J., Wang, X., Huang, J., Wang, Z., Li, Y., Wang, X., He, Y., Gu, C., Liu, M., Ma, S., Wen, S., & Qiu, P. (2022). Microhaplotype and Y-SNP/STR (my): A novel MPS-based system for genotype pattern recognition in two-person DNA mixtures. *Forensic Science International: Genetics, 59*. https://doi.org/10.1016/j.fsigen.2022.102705

Fekete, A., & Schmitt-Kopplin, P. (2007). Capillary electrophoresis. In *Food toxicants analysis* (pp. 561–597). Elsevier. https://doi.org/10.1016/B978-044452843-8/50016-X

Gandotra, N., Speed, W. C., Qin, W., Tang, Y., Pakstis, A. J., Kidd, K. K., & Scharfe, C. (2020). Validation of novel forensic DNA markers using multiplex microhaplotype sequencing. *Forensic Science International: Genetics, 47*. https://doi.org/10.1016/j.fsigen.2020.102275

Hou, R., Yang, Z. X., Li, M. H., & Xiao, H. S. (2013). Impact of the next-generation sequencing data depth on various biological result inferences. In *Science China life sciences* (pp. 104–109). Science in China Press. https://doi.org/10.1007/s11427-013-4441-0

Jian, H., Wang, L., Lv, M., Tan, Y., Zhang, R., Qu, S., Wang, J., Zha, L., Zhang, L., & Liang, W. (2021). A novel SNP-STR system based on a capillary electrophoresis platform. *Frontiers in Genetics, 12*. https://doi.org/10.3389/fgene.2021.636821

Kidd, K. K., & Pakstis, A. J. (2022). State of the art for microhaplotypes. *Genes, 13*(8). https://doi.org/10.3390/genes13081322

Kidd, K. K., Pakstis, A. J., Gandotra, N., Scharfe, C., & Podini, D. (2022). A multipurpose panel of microhaplotypes for use with STR markers in casework. *Forensic Science International: Genetics, 60*. https://doi.org/10.1016/j.fsigen.2022.102729

Kidd, K. K., Pakstis, A. J., Speed, W. C., Lagace, R., Wootton, S., & Chang, J. (2018). Selecting microhaplotypes optimized for different purposes. *Electrophoresis, 39*(21), 2815–2823. https://doi.org/10.1002/elps.201800092

Kidd, K. K., & Speed, W. C. (2015). Criteria for selecting microhaplotypes: Mixture detection and deconvolution. *Investigative Genetics, 6*(1). https://doi.org/10.1186/s13323-014-0018-3

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics, 25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Martinez-Martin, N., & Magnus, D. (2019). Privacy and ethical challenges in next-generation sequencing. *Expert Review of Precision Medicine and Drug Development* (Vol. 4,(2), 95–104. https://doi.org/10.1080/23808993.2019.1599685

National DNA Database Ethics Group. (2017). *ETHICAL dimensions of the application of next generation sequencing technologies to criminal investigations author*. National DNA Database Ethics Group.

Oldoni, F., Bader, D., Fantinato, C., Wootton, S. C., Lagacé, R., Kidd, K. K., & Podini, D. (2020). A sequence-based 74plex microhaplotype assay for analysis of forensic DNA mixtures. *Forensic Science International: Genetics, 49*, 102367. https://doi.org/10.1016/j.fsigen.2020.102367

Oldoni, F., Kidd, K. K., & Podini, D. (2019). Microhaplotypes in forensic genetics. *Forensic Science International: Genetics, 38*, 54–69. https://doi.org/10.1016/j.fsigen.2018.09.009

Pang, J. B., Rao, M., Chen, Q. F., Ji, A. Q., Zhang, C., Kang, K. L., Wu, H., Ye, J., Nie, S. J., & Wang, L. (2020). A 124-plex microhaplotype panel based on next-generation sequencing developed for forensic applications. *Scientific Reports, 10*(1). https://doi.org/10.1038/s41598-020-58980-x

Staadig, A., & Tillmar, A. (2021). Evaluation of microhaplotypes in forensic kinship analysis from a Swedish population perspective. *International Journal of Legal Medicine, 135*, 1151–1160. https://doi.org/10.1007/s00414-021-02509-y/Published

Standage, D. S., & Mitchell, R. N. (2020). MicroHapDB: A portable and extensible database of all published microhaplotype marker and frequency data. *Frontiers in Genetics, 11*. https://doi.org/10.3389/fgene.2020.00781

Tillmar, A. O., & Phillips, C. (2017). Evaluation of the impact of genetic linkage in forensic identity and relationship testing for expanded DNA marker sets. *Forensic Science International: Genetics, 26*, 58–65. https://doi.org/10.1016/j.fsigen.2016.10.007

Turchi, C., Melchionda, F., Pesaresi, M., & Tagliabracci, A. (2019). Evaluation of a microhaplotypes panel for forensic genetics using massive parallel sequencing technology. *Forensic Science International: Genetics, 41*, 120–127. https://doi.org/10.1016/j.fsigen.2019.04.009

Turchi, C., Pesaresi, M., & Tagliabracci, A. (2017). A microhaplotypes panel for forensic genetics using massive parallel sequencing. *Forensic Science International: Genetics Supplement Series, 6*, e117–e118. https://doi.org/10.1016/j.fsigss.2017.09.035

Wang, H., Zhu, J., Zhou, N., Jiang, Y., Wang, L., He, W., Peng, D., Su, Q., Mao, J., Chen, D., Liang, W., & Zhang, L. (2015). NGS technology makes microhaplotype a potential forensic marker. *Forensic Science International: Genetics Supplement Series, 5*, e233–e234. https://doi.org/10.1016/j.fsigss.2015.09.093

Yang, J., Chen, J., Ji, Q., Yu, Y., Li, K., Kong, X., Xie, S., Zhan, W., Mao, Z., Yu, Y., Li, D., Chen, P., & Chen, F. (2022). A highly polymorphic panel of 40-plex microhaplotypes for the Chinese Han population and its application in estimating the number of contributors in DNA mixtures. *Forensic Science International: Genetics, 56*. https://doi.org/10.1016/j.fsigen.2021.102600

Zhu, J., Lv, M., Zhou, N., Chen, D., Jiang, Y., Wang, L., He, W., Peng, D., Li, Z., Qu, S., Wang, Y., Wang, H., Luo, H., An, G., Liang, W., & Zhang, L. (2019). Genotyping polymorphic microhaplotype markers through the Illumina® MiSeq platform for forensics. *Forensic Science International: Genetics, 39*, 1–7. https://doi.org/10.1016/j.fsigen.2018.11.005

This page intentionally left blank

CHAPTER 8

Tools and techniques of using NGS platforms in forensic population genetic studies

Bofeng Zhu[1] **and Chunmei Shen**[2]

[1]Guangzhou Key Laboratory of Forensic Multi-Omics for Precision Identification, School of Forensic Medicine, Southern Medical University, Guangzhou, Guangdong, China; [2]Department of Laboratory Medicine, Nanfang Hospital, Southern Medical University, Guangzhou, Guangdong, China

Overview of the concept of population study

Population study, also referred to as population genetics study, constitutes a subfield of genetics that employs mathematical and statistical methods to study the variations of gene frequencies and genotype frequencies in biological populations, scrutinizes the population genetic structures and relationships, and explores the population genetic backgrounds. Moreover, population studies, especially population genetics studies, play a pivotal role in common forensic genetics applications, including individual identification, paternity testing, and ancestry informative inference. Hereby, we generally focus on the systematic studies of population genetics in forensic genetics and introduce the basic concepts of a population study, the theoretical basis of genetics, the relevant parameters of population genetics, as well as the corresponding forensic genetics applications.

The concept of population study

The term "population" generally refers to all individuals of the same species, such as all living human beings on the earth. The "population" mentioned in this chapter specifically refers to the human population of the context of forensic applications. The gene pool is defined as the entire collection of genes shared by individuals within a freely mating population. The genetic structure of a population is determined by the frequencies of various genes and the distributions of genotypes generated from different mating patterns. Each individual in the population inherits alleles from their biological parents and subsequently passes these alleles on to their offspring. Frequencies of genes and genotypes in the population are influenced by various evolutionary forces as the genes and genotypes pass down through generations and eventually lead to variations of allele polymorphisms in the gene pool. The evolutionary forces which affect the genetic diversity of the gene pool include mutation, gene drift, natural selection, environmental diversity, migration, and nonrandom mating pattern, etc.

Next Generation Sequencing (NGS) Technology in DNA Analysis
ISBN 978-0-323-99144-5, https://doi.org/10.1016/B978-0-323-99144-5.00026-3
© 2024 Elsevier Inc.
All rights reserved.

In order to investigate the distributions of gene frequencies and genotype frequencies among populations, population study applying mathematical and statistical methods is able to investigate the genetic divergence of gene frequencies and genotype frequencies among and within the populations. Exploring the interrelations between genetic structure variations and factors of evolutionary forces is of great importance in population genetics; the latter comprises environmental selection, genetic mutation, migration, and genetic drift, etc. The subjects of forensic population study are Mendelian populations, the relatively larger populations of the same species which propagate sexually, where individuals therein mate randomly and follow Mendel's laws of inheritance.

Genetic theories of population study

Population genetics is essentially a combination of genetics and evolutionary theories. In terms of genetics, classical population genetics focuses on fundamental principles like Mendel's laws of segregation and independent assortment, along with Morgan's law of linkage and crossing-over. This concentration aims to track dynamic population changes to make more precise predictions about specific population developments. In terms of evolutionary theories, scientists like R. A. Fisher, J. B. S. Haldane, and Sewall Wright developed and refined the "modern evolutionary synthesis" on the basis of Darwinism theory and mathematical statistics in the early 20th century (Servedio et al. 2014). After integrating with mathematical statistical models such as the Hardy-Weinberg equilibrium law and the Wright—Fisher model, classical population genetics has been perfected by eventually transforming into an organic unity of genetics and statistics. Since the dawn of molecular biology in the mid-to-late 20th century, scientists have found a myriad of genetic markers with different levels of polymorphisms which cannot be explained by current evolutionary theories. Therefore, Kimura (1968) introduced the "neutral theory of molecular evolution", which seems to be the counterpart of modern evolutionary synthesis, and is developed based on the theory of natural selection. However, these two views have dialectically compensated for each other's theoretical weakness and collectively formed the basis of modern population genetics.

Forensic applications of population study

Forensic genetics, as an indispensable branch of forensic science, has been rapidly developed and widely used since its inception. With the continuous expansion of its applications and scope of examinations, it has already become one of the most common and effective technologies for criminal/case investigations. However, the rise of intelligent criminal means has led to a higher proportion of high-tech crimes which puts forward more stringent demands for in-depth mining of biological evidence to tackle forensic-related issues such as individual identification, paternity testing, and ancestry inference. Forensic population studies mainly focus on exploring genetic markers with polymorphisms in certain populations, further calculating a series of forensic parameters of genetic

markers via population genetic structure information of allele frequencies and genotype frequencies, and eventually constructing genetic marker systems which are qualified to perform forensic individual identifications and paternity tests in different populations. In addition, forensic population studies regularly survey the genetic background of the target population and its genetic relationships with other populations by analyzing population genetic structure data with statistical methods so as to obtain genetic information like the allele and genotype frequencies of the targeted population and lay the foundation for calculating forensic parameters such as likelihood ratio and paternity index (PI) for the forensic individual identification, paternity test, etc.

Forensic genetics study requires massive data of genetic markers in different populations as a basis. Currently, plenty of population studies have been carried out on both the traditional capillary electrophoresis (CE) platform and the novel next-generation sequencing (NGS) platform. As a mainstream platform in current criminal/case investigations, the CE platform detects the genotypes of targeted loci based on the differences in fluorescent dye markers and allelic fragment lengths of loci and accordingly obtains the length polymorphism information of genetic markers. Compared with the CE platform, the second-generation sequencing platform is not only able to access the length polymorphism information of conventional genetic markers but also able to precisely read the sequence information, resulting in a more comprehensive acquisition of the genetic information of targeted markers and a more comprehensive evaluation of the forensic efficacies of different genetic marker systems. There are three main research topics in forensic population studies based on the NGS platform which still remain to be studied nowadays. Firstly, how to fully take advantage of the high-throughput NGS platform and apply more genetic markers with more different types for analyzing the genetic background of the targeted population, and also to provide a simultaneous evaluation of the forensic utility of multiple genetic markers as well as the related population genetics data; Secondly, how to explore the genetic structures, backgrounds, and relationships among different populations more effectively based on the comparative analysis of genetic data obtained from various populations; thirdly, how to investigate the independence of genetic markers with one or more types in the same or different panels based on the DNA profiling results from multiple genetic marker systems; survey the feasibility and efficacy of the combined application of multiple panels with different types of genetic markers; and also develop a statistical method for analyzing the forensic applicability of the multiplex systems using one or more types of genetic markers based on mass data.

Forensic parameters of population genetics

Candidate molecular genetic markers or detection systems consisting of genetic markers need to be evaluated for system performance before they are used in forensic practice. The evaluation mainly includes three interrelated aspects: one is to evaluate the

population applicability of candidate molecular genetic markers, that is, whether the related genetic markers conform to the Hardy-Weinberg equilibrium in the target population. Meanwhile it is in a state of linkage equilibrium in the target population; the second is to evaluate the polymorphisms of the candidate molecular genetic markers, that is, to obtain the genetic polymorphism parameters of genetic markers through a population survey. The purpose of this evaluation is to obtain molecular markers with higher levels of polymorphisms to achieve higher efficiency in forensic applications such as individual identification and paternity testing. The third aspect of the validation experiments is to identify the genetic relationships among populations and build statistical analysis methods for forensic applications such as kinship identification, and ancestry information source inference based on the constructed system.

Population applicability of candidate molecular genetic markers

Hardy-Weinberg equilibrium test

Hardy-Weinberg equilibrium, also known as the Hardy-Weinberg rule, is an important statistical test method in population genetics. This rule was discovered in 1908 by two scholars, Hardy, G. H., an English mathematician; and Wilhelm Weinberg, a German physician, this theorem was named after their surnames. The theorem asserts that under ideal conditions, the gene frequency and genotype frequency of a population remain unchanged from generation to generation, and it clarifies the effect of reproduction on the gene and genotype frequencies of a population. The significance of the Hardy-Weinberg equilibrium law has two main points: one is to reflect the relationship between gene frequency and genotype frequency; the second is to evaluate whether the result of the sampling survey is representative of the target population, and then to evaluate the reliability of the population research results. The Hardy-Weinberg equilibrium law requires five ideal conditions to be satisfied: an infinite population; random mating among individuals within the population; no mutation; no natural selection; and no migration. Only the candidate genetic markers are in Hardy-Weinberg equilibrium in the applied population, the statistical analyses and calculations of the likelihood ratio and PI based on the polymorphism data of these genetic markers can ensure the reliability and accuracy, which could further achieve the purposes of forensic individual identification and kinship identification. Therefore, for new candidate genetic markers, it is necessary to assess whether they are in Hardy-Weinberg equilibrium in the studied population.

Linkage and linkage disequilibrium

(1) Linkage

Linkage refers to the tendency of coinheritance of genes which are physically close to each other on the same chromosome during the meiotic stage of sexual reproduction. In the process of meiosis, the non-sister chromatids are crossed over, therefore, it can be considered that genetic markers with closer physical distances are more closely related

than those with longer physical distances. In other words, the closer two loci on a chromosome are, the lower the chance of recombination between them, and the more likely it is that they are co-inherited, such as the microhaplotype (MH) genetic marker, since it is in the range of 300 bp. A composite marker is composed of multiple SNPs, and different SNPs are inherited in linkage. In addition, genetic markers on different chromosomes are generally considered to be unlinked. For autosomes, the physical distance between genetic markers on the same chromosome is greater than 10 Mb, which should be unlinked. When genes or loci are randomly combined during genetic material transmission, they are in a state of linkage equilibrium. The unlinked loci in linkage equilibrium state are independent of each other, and the product rule can be applied to calculate the cumulative probability of the system (which includes multiple loci), and each independent genetic marker can be combined to calculate.

(2) Linkage disequilibrium

Linkage disequilibrium refers to a nonrandom combination phenomenon that the probability of simultaneous occurrence of alleles belonging to two or more loci is different from the expected probability of random occurrence in a population. For example, the alleles of each locus of the human leukocyte antigen gene are not completely random to form haplotypes, and the observed frequencies of haplotypes in the population data are different from the expected frequencies. Loci in linkage disequilibrium might be located in different regions on the same chromosome or on different chromosomes. The influencing factors of linkage disequilibrium pairwise loci include natural selection, genetic drift, mutation, recombination, and linkage.

Linkage and linkage disequilibrium are both related and distinct. Genes linked to each other are not only related to the physical distances between chromosomes but also sometimes to the inheritance pattern of the disease, which can be tested in pedigree data. Linkage disequilibrium reflects the non-independent co-occurrence of alleles at different loci and describes associations between alleles which can be tested in pedigree and population genetic study data. When two genetic markers are in linkage or linkage disequilibrium, it is necessary to use haplotype frequencies and related parameters to describe and evaluate the genetic structure of the population.

Genetic markers located on sex chromosomes

Humans have a pair of sex chromosomes, XX for females and XY for males. There are also various types of polymorphic DNA genetic markers on sex chromosomes, such as short tandem repeats (STRs), SNPs, and InDels. Among them, the mother can pass the two X chromosomes to the offspring randomly, and the inheritance pattern is similar to that of the autosomes. When male sex chromosomes undergo meiosis, the pseudoautosomal region, constituting approximately 5% of the total length at both ends of the Y chromosome, can exchange and recombine with the homologous segment of the X

chromosome. The inheritance pattern of genetic material in this segment is similar to that of the autosomes. The other 95% of the Y chromosome segment (male-specific region, MSY) is different, when genetic material is passed on from the father, it is transmitted on to the son by haplotype, and the sequence located in the specific region of the father's X chromosome is also passed on by haplotype to the daughter. Due to the special transmission pattern of the genetic material of sex chromosomes, the genotypes of the sex chromosomal genetic markers are more advantageous in the forensic identification of special cases and complex kinship relationships.

Due to the uniqueness of inheritance patterns, sex chromosomal genetic markers are calculated differently from autosomal chromosome markers when testing for Hardy-Weinberg equilibrium and linkage disequilibrium. In the case of X-chromosome genetic markers, the Hardy-Weinberg test first examines whether the gene frequencies of these markers in different genders are not statistically different. Furthermore, for female individuals, if the X-genetic markers adhere to the Hardy-Weinberg rule, they are considered to be in Hardy-Weinberg equilibrium within the targeted population. The genetic marker of Y-chromosome-specific region is linked to inheritance and passed on to male offspring as a whole, which is identified as a Y-chromosome haplotype. The X-chromosome genetic markers are distributed on the 153 Mb X chromosome, and the density of the genetic markers on the X-chromosome is comparable to that of the autosomes. When the physical distance between the two X-chromosome genetic markers is relatively close, the possibility of linked inheritance should also be considered. Among them, clusters of closely linked genetic markers, called linkage groups, should be used as haplotypes for statistical computational analysis.

Polymorphism parameters of genetic markers

Alleles composed of different sequences among different individuals or between two different DNA copies of the same individual form a DNA polymorphism. According to the structural characteristics of DNA genetic markers, DNA polymorphisms can be divided into two types: length polymorphisms and sequence polymorphisms. Length polymorphism refers to a polymorphism composed of DNA fragment length differences among alleles at the same locus. Currently, STR and InDel are widely used genetic markers of length polymorphisms. A sequence polymorphism marker refers to a polymorphism consisted of one or more base differences in the DNA sequences of different individuals or between two DNA copies at the same locus of the same individual. SNPs are one of the most common genetic markers of sequence polymorphisms and exist on both chromosomes and mitochondrial DNA.

Common parameters for evaluating genetic marker polymorphism include gene frequency, genotype frequency, haplotype frequency, heterozygosity, and polymorphism information content (PIC), etc. The polymorphism parameters of each marker can be obtained through a population survey, and the surveyed population size must be large

enough. According to the recommendations of the International Society of Forensic Genetics (Gusmão et al., 2017), for population genetic studies based on autosomal STR (A-STR), X chromosome STR (X-STR), autosomal SNP/InDel (A-SNP/InDel), and X chromosome SNP/InDel (X-SNP/InDel), the number of samples should be more than 500; Y chromosome STR (Y-STR) must be more than 400 individuals; Y chromosome SNP/InDel (Y-SNP/InDel) must be more than 300 individuals; mitochondrial DNA (mtDNA) control region sequence based on Sanger sequencing (Full mtDNA CR_Sanger) samples should be more than 200; mitochondrial full sequence (Full mtDNA molecule_Sanger) sequencing samples based on Sanger sequencing should be more than 100; mtDNA coding region SNP (mtDNA_SNPs in coding region) detection should be more than 200 samples; population studies based on NGS platforms should be greater than 50 individuals.

Allele frequency and genotype frequency

A gene at a specific position on a chromosome is called a locus, and the genes which are located in the same position of a pair of homologous chromosomes with differences in the primary structure of DNA, are called alleles. In a population, there are three or more possible alleles at one locus, which are called multiple alleles, such as the STR locus and the ABO blood group gene. When there are only two alleles at a locus in the population, it is considered a biallelic locus, such as SNP and InDel.

Allele frequency refers to the ratio of the number of a certain allele at a particular locus to the total number of alleles at the locus, which is a basic measure for evaluating the genetic structure of a population. The sum of all the allele frequencies at a given locus adds up to 1. Genotype frequency refers to the ratio of the number of individuals with a specific genotype to the total number of individuals in a population, and the sum of all genotype frequencies at this locus is 1. The population genetic structure is represented by the frequency of all alleles and the frequency of various genotypes in the population. For a population, there are two or more alleles at a certain locus, and the frequency of each allele is more than 0.01, which means that the locus has a genetic polymorphism in that population.

Haplotype frequency

When two loci are associated or linked, the single locus is not independent, and the law of probability multiplication cannot be used to calculate parameters such as cumulative individual identification probability. In such cases, haplotype frequencies are needed to describe the genetic structure of the population or to calculate forensically relevant probabilities. A haplotype refers to the combination of alleles at linked loci, and the alleles constituting the haplotype are inherited in a linked manner. The haplotype of autosomal genetic markers needs to be determined by family survey, because when two linked loci are all heterozygous, it is impossible to judge whether two loci are linked and inherited

based solely on the results of individual typing. Maternally inherited mtDNA and paternally inherited Y-chromosome genetic markers can be used to determine haplotypes based on individual typing results and can be calculated directly using counting method.

Heterozygote (He) and polymorphism information content

Heterozygosity is a traditional genetic parameter that refers to the proportion of heterozygotes among all genotypes for a genetic marker in a population. The more alleles of a genetic marker, the closer the value of a single allele frequency is to the average of all allele frequencies of the genetic marker, and the proportion of heterozygotes in all genotypes is also higher. The higher the heterozygosity, the greater the application efficiency in forensic personal identification. Since males have only one X chromosome, only female individuals were considered when assessing heterozygosity for X chromosome genetic marker. Heterozygosity is usually expressed as H/h, where h is the observed value of heterozygosity. The formula is as follows:

$$h = \text{number of heterozygotes/total number of individuals} \tag{8.1}$$

H is the expected value of heterozygosity, and the calculation formula is as follow:

$$H = n/(n-1)\left(1 - \sum_{i=1}^{k} p_i^2\right) \tag{8.2}$$

In the formula, n is the number of samples, k is the number of allele, the p_i represents the frequency of the ith allele of a genetic marker.

PIC is a quantitative indicator for evaluating the usefulness of a polymorphic locus. The value of *PIC* can be calculated based on the number of alleles of a genetic marker and the distribution frequencies of its alleles. Similar to heterozygosity, a higher number of alleles at the candidate locus and a more balanced frequency distributions of all the alleles result in a greater PIC value. Generally considered, when $PIC > 0.5$, genetic marker is rich in genetic information; $0.25 < PIC < 0.5$, the genetic information content of the candidate marker is general; when $PIC < 0.25$, the genetic information content of the candidate marker is less. The formula for calculating the *PIC* of a genetic marker is as follow:

$$PIC = 1 - \sum_{i=1}^{n} p_i^2 - \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} 2p_i^2 p_j^2 \tag{8.3}$$

P_i and p_j indicate the frequency of the ith and jth alleles of a genetic marker, n is the number of allele.

Forensic application analysis based on population genetics

Data from population genetics are the basis for forensic applications such as individual identification, paternity testing, and ancestry inference, et al.

Individual identification

Comparing the genotypes of molecular markers obtained by the detection of biological sample left at the crime scene with the forensic DNA database is an important method to detect unknown individual in forensic investigation practice. Due to the limited number of the used genetic markers or insufficient information about genetic markers, there may be instances of false positive and false negative results in the comparison of the sample genotype with the database genotype. Random match probability (RMP) and power of discrimination (DP) are indicators to evaluate the accuracy of the conclusion of individual identification by DNA typing and help to judge the matching results. For newly developed candidate loci or detection systems, RMP and DP can be used to measure the efficacy of a genetic marker or detection system for individual identification.

1. Random match probability.

The probability of random match is a commonly used parameter in population genetics that expresses the probability that two individuals randomly selected from the population have the same genotype or haplotype. In forensic identification, the RMP can be explained that when the genotypes of the two samples are the same, the on-site samples are not left by the suspect but by a random individual. And RMP calculates the posibility of this occurrence in the above situation. The smaller the probability, the less likely the on-site inspection material is left by a random individual rather than a criminal suspect, indicating that the genotype matching of the on-site inspection material and the suspect sample is less likely to be a random event, which is more supportive of the hypothesis that the two genotypes came from the same person, and also supports the hypothesis that the on-site inspection material was left by the suspect. Human individual identification compares two samples by a series of genetic markers to determine whether they are from the same individual.

For a genetic marker with multiple alleles on an autosome, the formula for calculating the probability that two individuals are randomly selected in a population to have the same genotype is as follows:

$$RMP = \sum_{i=1}^{n} p_i^2 \tag{8.4}$$

P_i indicates the frequency of the ith allele n indicates the number of alleles of the genetic marker.

Combined *RMP* (*CRMP*) refers to the probability that two individuals are randomly selected from a population and their genotyping results of multiple genetic markers are the same. *CRMP* can be calculated as follows.

$$CRMP = \prod_{j=1}^{n} RMP_j \tag{8.5}$$

RMP_j indicates the *RMP* of the jth allele, and n indicates the number of genetic markers.

2. Discrimination power.

Individual discrimination power is another parameter corresponding to the probability of random matching, which refers to the probability that two individuals are randomly selected from the population and their genotypes are different. The ability of a polymorphic genetic marker to discriminate among random unrelated individuals can be assessed by this parameter. The greater the number of alleles of a polymorphic genetic marker and the more uniform the distributions of different alleles, the higher the individual identification ability of the genetic marker. DP is also an index to evaluate the effectiveness of genetic marker in identifying irrelevant individuals. The higher the DP, the stronger the genetic marker is in identifying irrelevant individuals. For a single genetic marker with multiple alleles in an autosomal chromosome, the formula for calculating DP is as follows:

$$DP = 1 - \sum_{i=1}^{n} p_i^2 = 1 - RMP \tag{8.6}$$

P_i indicates the frequency of the ith allele n indicates the allele number of this marker.

When multiple genetic markers are used in combination, it is expressed by the *DP* (*CDP*). Calculated as follows:

$$CDP = 1 - \prod_{j=1}^{n} RMP_j = 1 - \prod_{j=1}^{n} \left(1 - DP_j\right) \tag{8.7}$$

RMP_j indicates the *RMP* of the jth allele, DP_j indicates the *DP* of the jth allele, n indicates the number of markers.

Parentage testing

Parentage testing is based on the DNA typing results of genetic markers and is analyzed according to the laws of inheritance to determine the parentage relationship between the alleged father (AF) or the alleged mother (AM) and their offspring. Probability of exclusion (PE) and cumulative PE (CPE), mutation rate, PI, and combined PI (CPI) are important parameters in parentage testing application.

1. Probability of exclusion and cumulative probability of exclusion.

Nonpaternal exclusion probability refers to the probability that non-biological father of the child can be excluded by a genetic detection system. This value is used to assess the utility of a genetic detection system in paternity testing case. The higher the polymorphisms of the genetic markers in the detection systems, the stronger the power to exclude nonbiological father. Taking the commonly used multiple allelic STR as an example, the p_i indicates the frequency of ith allele of a marker, p_j indicates the frequency of jth allele, the ith allele and jth allele are different, n indicates the number of alleles, then the nonpaternity exclusion rate of this locus is:

$$PE = \sum_{i=1}^{n} p_i(1-p_i)^2 - 1/2\left[\sum_{i}^{n-1}\sum_{j=1}^{n} {p_i}^2\,{p_j}^2\left(4-3\,p_i-3\,p_j\right)\right] \tag{8.8}$$

In practical applications, it is often necessary to combine multiple loci in the identification system to determine the possibility of excluding paternity, that is, the cumulative nonpaternal exclusion rate (*CPE*). When two or more genetic markers are independent of each other, the *CPE* can be calculated by the product of the probability law. The formula for calculating *CPE* is:

$$CPE = 1-(1-PE_1)(1-PE_2)(1-PE_3)...(1-PE_n) = 1-\prod_{I=1}^{n}(1-PE_n) \tag{8.9}$$

PE_n is the PE value of the marker n.

2. Mutation rate.

DNA replication error or mis-repair during DNA damage repair can lead to mutations. Mutations which occur during meiosis will be passed on to the next generation along with the genetic material, resulting in a phenomenon that does not conform to the laws of inheritance between parent and offsprings. As the number of tested genetic markers increases, the potential for variants also increases. The high throughput of the *NGS* platform will surely lead to an increase in the number of genetic markers which can be detected at once. Reasonably and accurately assessing the mutation rates of various genetic markers used in forensic practice is also an important purpose of conducting population genetic research based on the NGS platform. The primary method for assessing mutation rates in forensic genetic markers is based on pedigree samples with at least 500 meiosis.

(3) Paternity index and combined paternity index.

The PI is a parameter for judging the power of genetic evidence in paternity testing and is the likelihood ratio (LR) of two conditional probabilities, the ratio of hypothesis 1 "the probability (X) of a man with alleged father (AF) genetic marker genotypes is the biological father of the child" to hypothesis 2 "the probability (Y) that a random man is the biological father of the child". Calculated as follows:

$$PI = LR = \frac{X}{Y} \tag{8.10}$$

When multiple genetic markers are used for parentage determination and multiple genetic markers should be independent of each other, the *CPI* can be obtained by multiplying the *PI* of multiple markers. The formula for calculating the cumulative *PI* of n genetic markers is as follows:

$$CPI = PI_1 \times PI_2 \times PI_3 \times \cdots \times PI_n \tag{8.11}$$

PI_n is the *PI* value of the nth marker.

Inference of ancestry informative

For the investigation of suspect when the database comparison fails, inferring ancestry is one of the most effective measures to help lock the suspect and narrow the scope of the investigation. Through the analysis of ancestry informative markers with large differences in gene frequencies among different populations combined with the results of multiple population genetic parameters, the ancestry information of an unknown individual can be determined. A precise inference of the ancestry origin of the biological material at the scene can help narrow the scope of the investigation and improve the speed of solving case. The main parameters involved in the application of ancestry information inference are δ value and F_{ST} value.

1. δ value.

δ value represents the difference between the absolute allele frequencies of the same allele in two different regions or ethnic groups. The greater the difference in the allele frequency of the locus in the two populations, the larger the δ value. When screening ancestral informative loci, the δ value is one of the important reference indicators. When the δ value of two populations is larger than 0.5, the genetic marker is defined as the specific marker (population-specific markers) of the two target populations, which can be considered as the ideal genetic marker for ancestry-informative inference for the two target populations.

2. F_{ST} value.

Due to the differences in genetic structure and genetic background, the distributions of alleles and genotypes between different populations are quite different. When performing population genetics analysis, the polymorphism parameters of a certain population cannot be generalized and applied to another group with larger differences. Therefore, when analyzing several different populations, it is first necessary to understand whether there are differences between populations and the extent of the differences. The F_{ST} value is also called the fixation index (F_{ST}) or the genetic differentiation index between populations and is an important parameter to measure the degree of differentiation and genetic distance between populations.

Due to the differences and specificities of different molecular genetic markers in geographical distribution and population distribution, the thresholds of F_{ST} values are different for different purposes in different research projects when screening genetic markers. For example, for a genetic marker used for individual identification, the F_{ST} value should be as small as possible. The smaller the F_{ST} value, the smaller the difference in gene frequency of the genetic marker between different populations, and it can be used more widely for individual identification of multiple populations. For a genetic marker used for ancestry-informative inference, higher F_{ST} values between different populations are anticipated. And the larger the value, the greater the difference in gene frequency of the genetic marker between different populations, which can be

used to distinguish different populations. According to different target populations, different types of genetic markers, different numbers of genetic markers, and different application purposes, the researchers proposed different F_{ST} cutoff reference values.

Commonly used markers of population study on NGS platform

At present, DNA genetic markers in forensic population genetic research based on the different NGS platforms include length polymorphism genetic markers (e.g., STR), sequence polymorphism genetic markers (e.g., SNP), mitochondrial DNA, MH genetic markers, and compound genetic markers (e.g., InDel-STR, InDel-SNP, SNP-STR). These markers are mainly used for forensic individual identification, paternity testing, ancestry-informative inference, etc.

Short tandem repeat

Currently, STR is the most widely used genetic marker in forensic genetics, and it is also one of the important genetic markers for population genetic research based on NGS platform. Increasing the number of STR loci is an effective strategy to decrease the probability of random matching. However, due to the limitations of fluorescence types and the length of amplified fragments, the traditional CE platform has a limited number of detectable loci. The high-throughput NGS platform is not limited by the number of loci. Also, because the detection amplicon of the NGS platform is smaller, it is more suitable for trace and degraded samples. In addition, the NGS platform can obtain the sequence information of the core region and flanking region in addition to the detection of the repeat number of the core region of the STR, so that the sequence differences in these regions can be found and the polymorphism of the locus can be further improved. Katherine Butler Gettings et al. used 22 autosomal STR loci to sequence 183 unrelated African Americans, Caucasians, and Hispanics based on NGS, and found that 6 loci (D12S391, D2S1338, D21S11, D8S1179, vWA, and D3S1358) have more than twice as many alleles on sequence polymorphisms as alleles on length polymorphisms (Gettings et al., 2016). The CODIS STR loci are widely used in intra-laboratory-conducted systems and commercial kits because of their high polymorphisms. According to literature reports, in different populations, D3S1358, D5S818, D7S820, D8S1179, D13S317, D18S51, D21S11, and vWA and other loci with more complex core regions were found to have more variations, the D21S11 showed the largest increase in the number of alleles on sequence polymorphisms comparing with length polymorphisms, which was found among many populations (Gettings et al., 2016; Kwon et al., 2016; Novroski et al., 2016; Warshauer et al., 2015; Zhao et al., 2015). Wang et al. used the SeqTypeR24 CASE kit to sequence and analyze 291 Beijing Han individuals and found a total of 234 length

polymorphism-based alleles and 356 sequence polymorphism-based alleles. And 22 new sequence types in the core repeat regions were found at 9 STRs (D12S391, D1S1679, D21S11, D2S1338, D31358, D5S818, D6S474, FGA, vWA) (Chen et al., 2021). In addition, the researchers applied other commercial or self-built NGS-STR systems, indicating that the sequence-characterized STR typing obtained by the NGS platform has higher polymorphism characteristics, which can achieve higher individual identification accuracy (Barrio et al., 2019; Chen et al., 2021; Dai et al., 2019; Dash et al., 2021; Houston et al., 2018; Huszar et al., 2018; Kim et al., 2017; Lee et al., 2021; Liu et al., 2020; Miao et al., 2021; Silva et al., 2020; Wang et al., 2020; Warshauer et al., 2015; Wu et al., 2020; Zhang et al., 2018; Zhao et al., 2015, 2016). The high-throughput NGS platform allows a composite amplification system to contain more genetic markers and different types of markers. The NGS ForenSeq DNA Signature Prep Kit is a typical multi-marker composite detection system that can simultaneously achieve several forensic application goals. A number of studies based on the NGS ForenSeq DNA Signature Prep Kit in Chinese Han (Li et al., 2020), Chinese Tibetan (Li et al., 2021; Peng et al., 2020), Chinese Li (Fan et al., 2021), Chinese Hui group (He et al., 2021), France (Delest et al., 2020), White British, British Chinese, South Asian, West African and North East African (Devesse et al., 2020), Danish (Hussing et al., 2019), Saudi Arabian (Khubrani et al., 2019), US Caucasian, Hispanic, African American, and East Asian Chinese (Novroski et al., 2016), Yavapai Native American (Wendt et al., 2017) and Korean (Kim et al., 2018) and other populations on 26 A-STRs, 24 Y-STRs, and 6 X-STRs have shown that the heterozygosity, DP, PIC, RMP of STR typing technique based on sequence polymorphism analysis were larger than those of STR typing technique based on fragment length polymorphism detection. The cumulative RMP of 26 A-STR loci on sequence polymorphisms is over than 700 times higher than that of length-based loci alone (Devesse et al., 2020), which greatly enhances the polymorphic information contents of STR loci and improves the efficiencies of forensic applications such as individual identification and paternity testing. In addition, ancestry-informative inference is also an applied aspect of STR genetic markers. Based on the NGS platform, Hirak Ranjan Dash et al. conducted a study with 31 STRs on a central Indian population, showing that the central Indian population has relatively close genetic relationships with other Indian populations. Based on the genotyping results of these 31 STRs, they also observed that other reference populations clustered with each other according to geographical distances, and the genetic relationship results of populations showed strong correlation with geographical locations (Dash et al., 2021).

Single nucleotide polymorphism

SNPs are single-base polymorphisms in DNA sequences, such as substitutions, insertions, and deletions. SNP genetic markers are widely distributed in the human genome,

Table 8.1 STR studies on NGS platforms.

No.	Machines	Kits	Number of markers	Details of markers	Populations	Sample numbers	References	Years
1	MiSeq	Custom primers	28	28 Y-STRs	African American, Caucasian, and Hispanic	41	Warshauer et al. (2015)	2015
2	Ion torrent PGM	Custom primers	13	13 Y-STRs	Chinese Han population	42	Zhao et al. (2015)	2015
3	MiSeq	PowerSeq Auto System	24	22 A-STRs, 1 Y-STR, and Amelogenin	African American, Caucasian, and Hispanic	183	Gettings et al. (2016)	2016
4	MiSeq	A simplified PCR-based library preparation method	23	23 Y-STRs	Korean	250	Kwon et al. (2016)	2016
5	MiSeq FGx	ForenSeq DNA Signature prep Kit	58	27 A-STRs, 7 X-STRs and 24 Y- STRs	US Caucasian, Hispanic, African American, and East Asian Chinese	777	Novroski et al. (2016)	2016
6	Ion torrent PGM	Custom primers	10	10 A-STRs	Chinese Han	165	Zhao et al. (2016)	2016
7	MiSeq	Custom primers	23	23 A-STRs	Korean	250	Kim et al. (2017)	2017
8	MiSeq FGx	ForenSeq DNA Signature prep Kit	232	27 A-STRs, 7 X-STRs, and 25 Y- STRs, Amelogenin, and 94 iiSNPs, 56 aiSNPs, and 22 piSNPs	Yavapai Native American	62	Wendt et al. (2017)	2017
9	Ion torrent S5	Custom primers	12	12 A-STRs	Cannabis sativa	16	Houston et al. (2018)	2018

Continued

Table 8.1 STR studies on NGS platforms.—cont'd

No.	Machines	Kits	Number of markers	Details of markers	Populations	Sample numbers	References	Years
10	MiSeq FGx	PowerSeq Auto/ Mito/Y System kit	23	23 Y-STRs	multi-groups	100	Huszar et al. (2018)	2018
11	MiSeq FGx	MiSeqFGx forensic signature kit	27	27 A-STRs	Korean	209	Kim et al. (2018)	2018
12	Ion torrent PGM	Custom primers	34	34 A-STRs	Chinese Han	200	Zhang et al. (2018)	2018
13	Ion S5	Precision ID GlobalFiler NGS STR panel v2	31	31 A-STRs	Spanish individuals	496	Barrio et al. (2019)	2019
14	Ion torrent PGM	SeqType R16	16	16 A-STRs	Chinese Han	593	Dai et al. (2019)	2019
15	MiSeq FGx	ForenSeq DNA Signature prep kit	56	26 A-STRs, 24 Y-STRs, and 6 X-STRs	Danish	363	Hussing et al. (2019)	2019
16	MiSeq FGx	ForenSeq DNA Signature prep kit	118	27 A-STRs and 91 iiSNPs	Saudi Arabian	89	Khubrani et al. (2019)	2019
17	MiSeq FGx	ForenSeq DNA Signature prep kit	27	27 A-STRs	multi-groups	350	Aalbers et al. (2020)	2020
18	MiSeq FGx	ForenSeq DNA Signature prep kit	154	27 A-STRs, 7 X-STRs, and 25 Y- STRs, Amelogenin, and 94 iiSNPs	French	169	Delest et al. (2020)	2020
19	MiSeq FGx	ForenSeq DNA Signature prep kit	26	26 A-STRs	White British, British Chinese, South Asian, West African and North East African	1007	Devesse et al. (2020)	2020

20	MiSeq FGx	ForenSeq DNA Signature prep kit	56	26 A-STRs, 24 Y-STRs, and 6 X-STRs	Chinese Han	750	Li et al. (2020)	2020
21	MiSeq FGx	Custom primers	42	42 A-STRs	Chinese Hebei Han	58	Liu et al. (2020)	2020
22	MiSeq FGx	ForenSeq DNA Signature prep kit	154	27 A-STRs, 7 X-STRs, and 25 Y- STRs, Amelogenin, and 94 iiSNPs	Chinese Tibetan	107	Peng et al. (2020)	2020
23	MiSeq	PowerSeq Auto/Y system prototype kit	24	23 Y-STRs + SE33	US population	786	Silva et al. (2020)	2020
24	Ion S5XL	Precision ID GlobalFiler NGS STR panel v2	30	30 A-STRs	Chinese Tibetan and Han	115	Wang et al. (2020)	2020
25	MiSeq FGx	Custom primers	29	28 A-STRs loci + Amelogenin	Chinese Yunnan Bai	203	Wu et al. (2020)	2020
26	Ion torrent PGM	SeqTypeR24 CASE kit	23	23 A-STRs	Chinese Beijing Han	291	Chen et al. (2021)	2021
27	Ion S5	Precision ID GlobalFiler NGS STR Panel v2	31	31 A-STRs	Central Indian	138	Dash et al. (2021)	2021
28	MiSeq FGx	ForenSeq DNA Signature prep kit	154	27 A-STRs, 7 X-STRs, and 25 Y- STRs, Amelogenin, and 94 iiSNPs	Chinese Hainan Hlai	136	Fan et al. (2021)	2021
29	MGISEQ-2000	MPS-based Forensic Analysis System Multiplecues SetB Kit	134	52 STRs and 81 Y-STRs and one Y-InDel (M175)	Chinese Han	413	Fan et al. (2022)	2021

Continued

Table 8.1 STR studies on NGS platforms.—cont'd

No.	Machines	Kits	Number of markers	Details of markers	Populations	Sample numbers	References	Years
30	MiSeq FGx	Custom primers	23	23 A-STRs	Korean	101	Lee et al. (2021)	2021
31	MiSeq FGx	ForenSeq DNA Signature prep Kit	56	26 A-STRs, 24 Y-STRs, and 6 X-STRs	Chinese Tibetan	120	Li et al. (2021)	2021
32	BGISEQ-500	Custom primers	309	186 SNPs + 123 STRs	Chinese Han	73	Miao et al. (2021)	2021
33	MiSeq FGx	SifaMPS Panel	381	87 STRs and 294 SNPs	Chinese Han	50	Tao et al. (2021)	2021
34	MiSeq FGx	Custom primers	15	15 A-STRs	Chinese Northern Han	554	Zhang et al. (2018)	2018
35	Database	Custom primers	12	12 X-STRs	Filipino male individuals	143	Salvador et al.(2018) Chen et al. (2021)	2018
36	MiSeq FGx	ForenSeq DNA Signature prep Kit	58	27 A-STRs + 24 Y-STRs + 7 X-STRs	Chinese Hui	90		2021

so there are abundant loci for meeting forensic needs. Compared with STRs, SNPs have a lower mutation rate and higher stability in the population, which are suitable for ancestry-informative inference. However, as a biallelic genetic marker, SNPs have lower polymorphisms at single locus than STRs, and thus require the combined detection of a larger number of SNP loci to achieve the accuracy of individual identification. NGS has the characteristics of high throughput and accurate detection result, which can satisfy the requirements of simultaneous detection of a large number of SNPs and can also meet the needs of SNP typing with different purposes, which greatly improves the forensic efficiencies of SNPs in practice and also expands the application prospects of SNPs. According to the opinions of the 22nd International Conference on Forensic Genetics, the forensic applications of SNPs are currently mainly divided into individual identification SNPs, ancestry-informative SNPs (AISNPs), lineage-informative SNPs, and phenotype-informative SNPs (PISNPs). However, regardless of the type of SNP marker, its expected forensic efficacy has limited applicability in other populations (nontargeted design). For example, AISNPs which are effective for some populations may be ineffective in identifying other populations, so different AISNPs need to be screened for different populations. Zhang et al. (2017a,b) performed sequencing analyses of 273 and 231 SNP loci through the Ion Torrent PGM sequencing platform, proving that the platform and the genetic markers (SNPs) are of great value in individual identification and complex kinship identification, respectively. In recent years, more and more scientists have developed NGS-SNP systems to solve the problems of individual identification and biogeographic ancestry-informative inference. Li et al. (2017) screened 175 SNPs at the genome-wide level and evaluated the individual identification performance of the system in 54 reference groups in the world. The results denoted that the system had good individual identification efficiencies in these populations, and its average RMP can range from 4.77×10^{-71} to 1.06×10^{-64}. Guo et al. (2020) sequenced and analyzed 30 AISNPs in 127 Han individuals from Chinese Shaanxi province on the Illumina NextSeq 500 platform. And the target group and the 22 reference populations from the 1000 Genomes Project (1000 Genomes Release Phase 3) were combined to analyze the genetic background of the Chinese Shaanxi Han using population genetics statistical methods, such as STRUCTURE, F_{ST}, principal component analysis (PCA). And their results showed that Chinese Shaanxi Han population and East Asian populations have more similar ancestral components. Bulbul et al. (2018) analyzed 86 AISNP loci of 2568 individuals from 39 groups from Northeastern Africa, southwestern Asia, and Europe (EUR) to the Ural Mountains and found that these 86 AISNPs can effectively evaluate population genetic differences in target regions and might be used to inference the ancestry informative.

Based on the Ion Torrent PGM sequencing platform, Thermo Fisher has developed two commercial kits for individual identification and biogeographic ancestry informative inference which have received widespread attentions from researchers: the Precision ID

Identity Panel and Precision ID Ancestry Panel, respectively. The Precision ID Identity Panel consists of 90 autosomal SNPs and 34 Y-SNPs, while the Precision ID Ancestry Panel consists of 165 autosomal AISNPs. These two kits can be configured with automated library construction and template preparation workflows for easy operation. In addition, the small amplicon fragments designed in the kit are also more suitable for the analysis of forensic degradation samples. At present, studies have reported that the Precision ID Identity Panel has great potential for forensic individual identification in Chinese Tibetans, Hui group (Liu et al., 2018), northern and southern Han populations (Guo et al., 2016; Li et al., 2018), Brazilians (Avila et al., 2019), and Indians (Dash et al., 2022). Liu et al. (2018) found that the cumulative matching probability (CMP) values of 124 SNPs in Tibetan and Hui populations in China were 2.5880×10^{-33} and 4.6326×10^{-34}, respectively. And their CPE values were 0.999999386152271 and 0.999999696360182, respectively. A population genetic analysis based on the same 124 SNPs found that the Tibetan and Hui groups had the close genetic relationships with East Asian populations. Similarly, the current studies also explored the applications in biogeographic ancestry informative inference on the Precision ID Ancestry Panel in Danes, Somalis (Pereira et al., 2017), Chinese Kazaks (Xie et al., 2020), Chinese Han, Chinese Li, Chinese Gelao (He et al., 2021), and seven other different Asian populations (Lee et al., 2018). Xie et al. (2020) sequenced and analyzed 165 AISNP genotypes of 96 Chinese Kazak individuals, and found that the Chinese Kazak group has the mixed ancestral components of East Asian and European, with a ratio of about 62:37. What's more, the SNP-detecting systems designed by many scientists also have good performances in individual identification and ancestry-informative inference. For example, a system containing 1245 SNPs was constructed by Wu et al. (2019); a 165 Y-SNPs system was constructed by Wang et al. (2019), a 144 SNPs and 20 MHs system was constructed by Phillips et al. (2019); a 153 SNPs system was constructed by de la Puente et al., (2021); and a 448-plex SNPs system was constructed by Zhao et al. (2021). In addition, the phenotypic characteristics of individuals can also be predicted by analyzing population SNP genotypes. Salvoro et al. (2019) evaluated the role of four models (IrisPlex, Ruiz, Allwood, and Hart) in the eye color prediction of 296 Italians through 24 SNPs, respectively. Bulbul and Filoglu (2018) analyzed the ancestral informational components and phenotypic characteristics of populations in Southwest Asia, EUR, and other continents through 156 SNPs, and their results showed that the system could be used for ancestry-informative inference and prediction of the eye, hair, and skin colors of some regional populations.

The development of NGS technology provides an accurate and reliable method for in-depth study of the application of SNPs in forensic individual identification, biogeographic ancestry informative inference, and phenotype prediction. With the continuous innovation of NGS technology and the discovery of more high-discrimination SNPs, future forensic identification will be more accurate and efficient. Table 8.2.

Table 8.2 SNP studies on NGS platforms.

No.	Machines	Kits	Number of markers	Details of markers	Populations	Sample numbers	References	Years
1	Ion torrent PGM	Custom primers	231	231 SNPs	Chinese Han population	50	Zhang et al. (2017b)	2017
2	Ion torrent PGM	Custom primers	273	273 SNPs	Chinese Han	50	Zhang et al. (2017a)	2017
3	Ion torrent PGM	Custom primers	175	175 SNPs	Chinese	729	Li et al. (2017)	2017
4	Illumina Nextseq	Custom primers	30	30 SNPs	Chinese Han	127	Guo et al. (2020)	2020
5	Miseq	Custom primers	86	86 SNPs	39 reference populations	2568	Bulbul et al. (2018)	2018
6	Ion torrent PGM	Precision ID Identity Panel	124	124 SNPs (90 SNPs + 34 Y-SNPs)	Tibetan, Uygur, and Hui minority ethnicities of China	184	Liu et al. (2018)	2018
7	Ion torrent PGM	HID-Ion AmpliSeq Identity Panel	124	124 SNPs (90 SNPs + 34 Y-SNPs)	Southern and Northern Chinese Han	108	Guo et al. (2016)	2016
8	Ion torrent PGM	HID-Ion AmpliSeq Identity Panel	124	124 SNPs (90 SNPs+34 Y-SNPs)	Southern Chinese	125	Li et al. (2018)	2018
9	Ion torrent PGM	HID-Ion AmpliSeq Identity Panel	124	124 SNPs (90 SNPs + 34 Y-SNPs)	Brazilian regional populations	432	Avila et al. (2019)	2019
10	Ion torrent PGM	Precision ID Identity Panel	124	124 SNPs (90 SNPs + 34 Y-SNPs)	Indian	95	Dash et al. (2022)	2021

Continued

Table 8.2 SNP studies on NGS platforms.—cont'd

No.	Machines	Kits	Number of markers	Details of markers	Populations	Sample numbers	References	Years
11	Ion torrent PGM	Precision ID Ancestry Panel	165	165 SNPs	Danes and Somalis	240	Pereira et al. (2017)	2017
12	Ion torrent PGM	Precision ID Ancestry Panel	165	165 SNPs	Chinese Kazak	96	Xie et al. (2020)	2020
13	Ion torrent PGM	Precision ID Ancestry Panel	165	165 SNPs	Haikou Han, Qiongzhong Hlai and Daozhen Gelao of China	316	He et al. (2021)	2021
14	Ion S5XL	HID-Ion AmpliSeq Ancestry Panel	165	165 SNPs	Southern China, Beijing, Japan, Korea, Vietnam, Nepal, India and Pakistan	750	Lee et al. (2018)	2018
15	Illumina MiSeq and HiSeq	Custom primers	1245	1245 SNPs	Chinese	210	Wu et al. (2019)	2019
16	Ion torrent PGM	Custom primers	165	165 Y-SNPs	Unknown	53	Wang et al. (2019)	2019
17	Ion torrent PGM	Custom primers	164	144 SNPs + 20 microhaplotypes	31 populations from 1 KG, SGDP and MAPlex genotyping	2333	Phillips et al. (2019)	2019
18	Miseq FGx	Custom primers	153	153 SNPs	Individuals from Middle East, North, and East African regions	167	de la Puente et al. (2021)	2021

19	BGISEQ-500RS	Custom primers	448	448-Plex SNP	Chinese Han	142	Zhao et al. (2021)	2021
20	Ion torrent PGM	Custom primers	156	156 SNPs	Turkey, Turkish Cypriots, Druze, Somali, Zaramo, Afghanistan, Spain	11	Bulbul and Filoglu (2018)	2018
21	Ion torrent PGM	Custom NGS-based panel	24	24 SNPs	Italian	296	Salvoro et al. (2019)	2019
22	Ion torrent PGM	Precision ID Ancestry Panel	165	165 SNPs	Unknown	95	Al-Asfi et al. (2018)	2018
23	Miseq FGx	ForenSeq DNA Signature prep Kit	165	165 SNPs	African American, East Asian,US Caucasian, Southwest US Hispanic	714	King et al. (2018)	2018

Mitochondrial DNA

MtDNA is a maternally inherited extra-nuclear genetic material. The mtDNA molecule is a circular double-stranded DNA with a length of 16,569 bp, which can be divided into two parts: the coding region and the noncoding region (also called the control region). Among them, the coding region is relatively conserved, and a total of 37 genes encode various RNAs and proteins. The sequence length of the control region (CR) of mtDNA is 1125 bp, and the mutation rate of the control region is relatively high. The control region is usually divided into three hypervariable regions: hypervariable region I (HVI), hypervariable region II (HVII), and hypervariable region III (HVIII). Due to the characteristics of multiple copies, high mutation rate, non-recombination, and maternal inheritance, mtDNA plays an important role in the identification of forensic matrilineal kinship, the identifications of trace and degraded samples, and population genetic studies. In 2014, based on the NGS platform (Roche 454), Mikkelsen et al. (2014) obtained the accurate point mutation genotypes of mtDNA HV I and HV II as previously detected with traditional CE sanger sequencing, single base extension (SNaPshot), and matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF). Subsequently, a number of large-scale population studies based on NGS platforms such as the Personal Genome Machine (PGM), Ion S5, Illumina MiSeq FGx, etc. confirmed that NGS can obtain highly reproducible and accurate genotyping of full mtDNA sequences (Chaitanya et al., 2015; Kim et al., 2018; Lopopolo et al., 2016; Park et al., 2017; Peck et al., 2016; Zhou et al., 2016). Accurate NGS-mtDNA typing of highly degraded samples, such as skeleton aged 50—70 years (Avila et al., 2022; Gorden et al., 2018), skeleton from the historical shipwreck of La Belle (Ambers et al., 2020), skeleton up to 8 kyas (Eduardoff et al., 2017) in multiple studies confirmed the suitability of mtDNA for forensically degraded samples. In addition, a large number of NGS-mtDNA-based population genetic studies of mother-child pairs and unrelated individuals have explored the mutation rate, heterogeneity, maternal inheritance pattern, and genetic relationships. The results have demonstrated that mtDNA exhibits a high degree of polymorphisms across diverse populations, making it a promising tool for individual identification and tracing matrilineal pedigrees (Ma et al., 2018; Park et al., 2017). Furthermore, a series of population studies have been carried out based on mtDNA hypervariable regions, control regions, and whole mtDNA genome to explore the genetic structures and evolutionary origins of different populations across continents from the perspective of maternal genetic material, providing supports for forensic ancestral informative inference. Among them, a population study based on NGS-mtDNA on Greenland showed that the Inuits of Greenland were descended from Thule, and factors such as geographic isolation and genetic drift had a significant impact on the genetic diversity of Greenlanders (Lopopolo et al., 2016). The Korean population analysis using the whole mtDNA sequence on the NGS platform showed that the Koreans have closer genetic relationships with the East Asian populations, suggesting that the Koreans have the same ancestral origin as the East Asians comparing with the individuals from other

continental populations (Park et al., 2017). A Sherpa-specific subhaplogroup A15c1 was found in a NGS-mtDNA population genetic study on the Chinese Muli Tibetan, suggesting that the Sherpa people may have originated from the Tibetan group, while the mitochondrial haplogroup of the Chinese Muli Tibetan group shared with the Han population, which may be due to intermarriage with each other (Wang et al., 2020). An NGS-mtDNA study of southern Brazil populations showed that in southern Brazil, there are populations which are farther away from the Brazilian populations but more genetically related to the European populations (Avila et al., 2022). In addition, EMPOP is a widely used forensic mtDNA database (https://empop.online). So far, EMPOP has collected 48,572 high-quality mtDNA microtypes, of which 46,963 cover HVS-I and HVS-II, 38,361 cover the control region, and 4289 are the whole mtDNA microtypes. Among the data collected by EMPOP, 19,605 mitotypes are from America, 13,721 are from EUR, 12,196 are from Asia, 2577 are from Africa (AFR), and 473 are from Australia and Oceania populations. Table 8.3 is a summary of forensic genetics in NGS-mtDNA population studies, including sequenced regions, study populations, population sizes, NGS platforms, and references.

Microhaplotype

A MH is a kind of genetic marker consisting of at least two single nucleotide polymorphisms (SNPs) within a certain range (300 bp) of DNA fragments. MH has several characteristics of no stutter peaks, abundant polymorphisms, and small amplicons, and also has the potential to be applied in forensic individual identification, mixture deconvolution, ancestral informative inference, and kinship testing. Because of these advantages, MH is an important type of genetic marker for population genetic research based on the NGS platform. Van der Gaag et al. (2018) developed a multiple amplification system containing 16 MHs (at least 4 SNPs in the range of 70 bp), which can accurately classify the 276 studied unrelated individuals into three great intercontinental origins namely EUR, AFR, and East Asia (EAS); additionally, this system confirmed the effectiveness of MHs in mixture deconvolution. In 2018, Peng et al. developed a multiple amplification system containing 25 MHs (amplicons less than 50 bp) based on the HiSeq X-10 platform (Chen et al., 2018). The results of population genetic study on unrelated individuals and artificially mixed samples of Chinese Han nationality showed that the sequencing depth of MHs has a significant correlation with the mixture ratio, and has the potential to deconvolve mixtures (Chen et al., 2018). Subsequently, a number of scientists based on MiSeq, Ion S5, and NextSeq 500 sequencing platforms developed a series systems containing 15 (Wang et al., 2022), 20 (Kureshi et al., 2020), 21 (Zou et al., 2022), 25 (Chen et al., 2019), 26 (Zhao et al., 2022), 30 (Wen et al., 2021), 40 (Yang et al., 2022), 59 (Kureshi et al., 2020), 90 (Pakstis et al., 2021), 140 (de la Puente et al., 2017), and 124 (Pang et al., 2020) MH loci. Based on the above systems, the researchers applied intercontinental populations (from AFR, EUR, EAS, South Asia, and the Americas), artificially simulated mixed samples, or degraded samples to confirm the effectiveness of individual

Table 8.3 Studies of mtDNA on NGS platforms.

No.	Machines	Regions	Populations	Sizes	References	Years
1	Roche 454 GS Junior	HV1 and HV2	Somalia and Dane	26	Mikkelsen et al. (2014)	2014
2	Ion torrent PGM	Whole mtDNA	European	20	Chaitanya et al. (2015)	2015
3	MiSeq	Whole mtDNA	Greenlandic	127	Lopopolo et al. (2016)	2016
4	MiSeq	Whole mtDNA	African American, U.S. Caucasian, and U.S. Hispanic	90	Peck et al. (2016)	2016
5	MiSeq	Whole mtDNA	Estonian	114	Stoljarova et al. (2016)	2016
6	Ion torrent PGM	Whole mtDNA	Chinese Northern Han	125	Zhou et al. (2016)	2016
7	Ion torrent PGM	mtDNA control region	DNA up to 8 kyrs of age	23	Eduardoff et al. (2017)	2017
8	Ion torrent PGM	Whole mtDNA	Korean	186	Park et al. (2017)	2017
9	MiSeq FGx	Whole mtDNA	21 nonprobative, degraded skeletal sample (aged 50–70 years)	21	Gorden et al. (2018)	2018
10	MiSeq	Whole mtDNA	mother-child pair of European decent	39	Holland et al. (2018)	2018
11	Ion torrent PGM	Whole mtDNA	Known innate heteroplasmy	5	Kim et al. (2018)	2018
12	Ion torrent PGM	Whole mtDNA	mother-child pairs of European decent	194	Ma et al. (2018)	2018
13	MiSeq FGx	Whole mtDNA	African American, US Caucasian, and US Hispanic	175	Peck et al. (2016)	2018
14	Ion torrent PGM	Whole mtDNA	Brazilian	96	Avila et al. (2019)	2019
15	Ion S5	Whole mtDNA	24 populations	526	Strobl et al. (2019)	2019
16	Ion S5	Whole mtDNA	shipwreck of La Belle	2	Ambers et al. (2020)	2020
17	Ion torrent PGM	Whole mtDNA	Romanian	43	Melchionda et al. (2020)	2020
18	MiSeq FGx	Whole mtDNA	five US metapopulations African American, Caucasian, Hispanic, Asian American, and Native American	1327	Taylor et al. (2020)	2020
19	Ion S5	Whole mtDNA	Chinese Tibetan	119	Wang et al. (2020)	2020
20	NextSeq 2000	Whole mtDNA	Southern Brazilian population	480	Avila et al. (2022)	2021
21	Ion S5	Whole mtDNA	Vietnam 52 age-old skeletal remain	50	Ta et al. (2021)	2021
22	MiSeq FGx	Whole mtDNA	African American, Caucasian, and US Hispanic	689	Wisner et al. (2021)	2021

identification, ancestral informative inference, degraded material identification, and mixture deconvolution, especially mixtures with large difference in proportion. The potential for secondary kinship analysis by MH markers has also been confirmed among the familiar population studies using the systems developed based on the NGS platform, which contained 13 (Zhu et al., 2019), 54, 59 (Wu et al., 2021), and 140 (de la Puente et al., 2017) MH markers (Sun et al., 2020; Wu et al., 2021). In addition, Chen et al. (2019) analyzed 10 MHs in the 1 KG database, showing that these 10 MHs can clearly distinguish the four intercontinental populations of AFR, EAS, South Asia, and EUR, and also indicating that these 10 MHs have the potential for ancestral informative inference. A system comprising 90 MH markers developed by Gandotra et al. (2020) has demonstrated good discrimination ability between African, Asian, and European ancestries. In addition, other several studies based on different continental populations also have further verified the efficiency of MH markers in forensic ancestry informative inference (Pakstis et al., 2021; Kidd et al., 2021; Wen et al., 2021; Zhao et al., 2022). The MH detecting system developed by Dan et al. based on NGS also included 10 MHs containing phenotypic SNPs (piSNPs) alongside 10 conventional highly polymorphic MH markers, allowing for the prediction of phenotypes (Wen et al., 2021). Based on the NGS platform, Jin et al. developed a complex detection system containing multiple types of markers (18 multi-allelic InDels, 27 MHs, and 36 Y-SNPs/InDels), showing good potential for individual identification, paternity testing, mixture deconvolution, ancestry informative inference, and Y haplogroup division (Jin et al., 2020). Table 8.4.

Compound marker

A compound marker refers to a new type of genetic marker formed by the combination of two different types of genetic markers, including several InDels, SNPs, and STRs within a range of less than 300 bp. Common compound markers include InDel-SNP, InDel-InDel, SNP-STR, InDel-STR, etc. The NGS platform takes into account the length and sequence information of the genetic markers and provides an ideal approach for the detection of composite markers. Based on the Illumina X-10 platform, Jin et al. developed a detection system for forensic individual identification containing 10 MHs and 13 InDel-SNPs. The CPD value of this system in Chinese Shaanxi Han population reaches 0.999999999994835, which indicates the potential of individual identification (Jin et al., 2020). The 23-plex system showed high polymorphism patterns in other 26 populations from five continents (AFR, America, EAS, South Asia, and EUR), suggesting that the system has a high potential for individual identification in major intercontinental populations around the world. In addition, phylogenetic analysis of five continental individuals based on 23-plex system loci showed that African populations were independently divided into a large clade, European and American populations were clustered into a large clade, and South Asian and East Asian populations clustered into a large clade. In addition, South Asia and EAS populations are further divided

Table 8.4 Studies of microhaplotypes on NGS platforms.

No.	Platforms	Number of markers	Details of markers	Populations	Sizes	Years	References
1	HiSeq X10	14	A-MHs	Chinese Han and 2 artificially prepared mixture samples	60	2018	Chen et al. (2018)
2	MiSeq	16	A-MHs	276 samples of 3 globally dispersed populations (5 Dutch, 3 Bhutanese, 2 Ghanese, 2 pygmy and 3 Amer- indian sample), 97 samples of 9 large families	373	2018	van der Gaag et al. (2018)
3	HiSeq X10	25	A-MHs	Chinese Han	60	2019	Chen et al. (2019)
4	Database	10	A-MHs	1 KG	2504	2019	Chen et al. (2019)
5	MiSeq	13	A-MHs	Chinese individuals	37	2019	Zhu et al. (2019)
6	MiSeq and Ion S5	118	107 A-MHs + 11 X-MHs	Cell line	5	2020	de la Puente et al. (2020)
7	MiSeq	90	A-MHs	7 African populations (Biaka, Chagga, Sandawe, Zaramo), Asia (taiwanese Chinese), individuals of mixed European ancestry (EuroAmericans), and an Eastern European population (Adygei)	156	2020	Gandotra et al. (2020)

8	MiSeq PE300	20	A-MHs	Chinese Han 50 unrelated individuals and 12 parent/child duo	74	2020	Kureshi et al. (2020)
9	MiSeq FGx	124	124 A-MHs+20 STRs	Chinese Han	256	2020	Pang et al. (2020)
10	MiSeq PE300	30	A-MHs	54 unrelated random Chinese Han +38 parent-child duos and 55 uncle/aunt/grandparent-child duos	94	2020	Sun et al. (2020)
11	MGISEQ-2000RS	129	48 AISNPs, 18 multi-allelic InDels, 27 microhaplotypes and 36 Y-SNPs/InDels	120 Shaanxi Han, 100 Mongolian and 115 Hui individuals	335	2021	Jin et al. (2022)
12		151	83 AISNPs + 68 A-MHs	95 population of Kidd Lab	3790	2021	Kidd et al. (2021)
13	MiSeq	90	90 A-MHs	16 populations of Kidd Lab	524	2021	Pakstis et al. (2021)
14	MiSeq	30	30 A-MHs including 10 phenotypic MHs	Northern Han Chinese	100	2021	Wen et al. (2021)
15	MiSeq FGx	54	54 A-MHs	Two donor families from the northern Han Chinese population containing 67 individuals	67	2021	Wu et al. (2021)
16	NextSeq 500	59	59 A-MHs	Han Chinese	187	2021	Wu et al. (2021)
17	Miseq PE300 platform	26	26 A-MHs, non binary SNPs with molecular extent sizes no longer than 60 bases	Northern Han Chinese	100	2021	Zhao et al. (2022)

Continued

Table 8.4 Studies of microhaplotypes on NGS platforms.—cont'd

No.	Platforms	Number of markers	Details of markers	Populations	Sizes	Years	References
18	A SNaPshot and phase workflow/Ion S5 XL system	21	21 A-MHs	10 Chinese populations (74 Chengdu Hans, 76 Dujiangyan Tibetans, 77 Muli Tibetans, 78 Xichang Yis, 78 Wuzhong Huis, 63 Zunyi Gelaos, 78 Hainan Lis, 80 Hainan Hans, 73 Ordos Mongolians and 87 tibet Sherpas)	764	2021	Zou et al. (2022)
19	Ion S5 XL System/ MinION nanopore sequencing	15	15 A-MHs	Chinese Han	70	2022	Wang et al. (2022)
20	HiSeq X-10	40	40 A-MHs	Chinese Han population	150	2022	Yang et al. (2022)

into two independent subclades, showing the 23-plex system's good ancestry informative inference ability and the potential to distinguish the ancestral origins of four intercontinental populations in AFR, EAS, South Asia, and EUR (Jin et al., 2020). Based on the NextSeq 500 platform, Jin et al. (2022) developed a complex amplification system containing 29 genetic markers (22 MHs and 7 compound markers); the heterozygosity values of the 29 markers in 26 reference populations and Chinese two populations (Kirgiz and Mongolian) were greater than 0.4. The CMP values of these markers ranged from $3.1928E^{-22}$ in the East Asian populations to $1.7023E^{-19}$ in the European populations. The CMP values were smaller than the values in previous 30 InDels and 35 InDels systems, suggesting the polymorphism of the 29-plex system and the efficacy of forensic applications such as individual identification and paternity testing. In addition, 29 markers showed large genetic differences in different continental populations, which is an accurate and effective forensic application system for distinguishing four continental populations in AFR, EAS, South Asia, and EUR (Jin et al., 2022) (Table 8.5).

Statistical analytical methods of population genetic study

Statistical analytical methods of population genetic study mainly include genetic differentiation index (F_{ST}) estimation, F test, multidimensional scaling (MDS) analysis, PCA, phylogenetic tree construction, and genetic mixed model prediction.

Genetic differentiation index (F_{ST})

The genetic differentiation index (F_{ST}) is a widely used index to measure the degree of genetic differentiations among populations. In those populations which conform to Hardy-Weinberg equilibrium, the larger the F_{ST} value between each other, the greater the difference, which is suitable for the comparison of diversity among subgroups. F_{ST} evolved from F-statistics. There are three main types of F-statistics (F_{IS}, F_{IF}, F_{ST}). F_{ST} is designed for the biallelic markers. If there are multiple alleles at the locus, it needs to be measured by the gene differentiation coefficient (G_{ST}). Assuming that there are

Table 8.5 Studies of compound markers on NGS platforms.

No.	Machines	Number of markers	Details of markers	Populations	Sizes	Years	References
1	NextSeq 500	29	22 MHs + 7 compound markers	112 Kazaks and 106 Mongolian	228	2020	Jin et al. (2020)
2	Illumina X- 10	23	10 MHs + 13 compound markers	Shaanxi Han population	141	2020	Jin et al. (2020)

populations, the relative size of the k-th population is w^k, the ith allele frequency in the k-th population is $q^k(i)$, and the observed heterozygote frequency is h^k, then the observed mean heterozygote frequency value in the whole population is HI, the mean expected heterozygote frequency of the population is HS, and the expected mean value of the heterozygote frequency of the entire population is HT, they are:

F_{IS} is the ratio of HI to HS reduction, or the mean inbreeding coefficient of the local populations.

F_{ST} is the ratio of HS to HT reduction, that is, the average inbreeding coefficient between related local populations.

F_{IT} is the ratio of HI to HT reduction, the average inbreeding coefficient for the entire populations.

The F_{ST} value ranges from 0 to 1, with a maximum value of 1 indicating complete differentiation between the two populations; and a minimum value of 0 indicating no differentiation between the populations. In actual research, when the F_{ST} value is 0—0.05, it indicates that the genetic differentiation between these two populations is very small and can be ignored; when it is 0.05—0.15, it indicates that there are moderate genetic differentiations between populations; when it is 0.15—0.25, there are large genetic differentiations between populations; when it is above 0.25, there is a large genetic differentiation between populations. The calculation of F_{ST} generally uses VCFtools software, genepop software, Arlequin software, and so on.

F test

The F test mainly includes three-group test and four-group test. The three-group test f3 is a powerful method for analyzing population genetic data and providing reliable evidence for population mixing events, even those which have occurred within the last few hundred generations. Firstly, the f3 (X, Y; Outgroup) test is usually used to analyze the shared gene drift between population X and population Y. A significant positive value indicates that after population X and population Y are separated from the other populations, gene exchange has occurred between the two populations (X and Y); then, we used the f3 (X, Y; and Z) test to analyze the admixture event of population Z; while significant negative value indicates that population Z is mixed by related populations of X and Y populations. The four-group test f4 can not only analyze population mixing but also provide discernible information on the direction of gene flow. Two types of f4 statistical tests are usually used in population studies: the symmetry f4 test and the affinity f4 test. The detection units of the symmetry f4 test include test group 1, test group 2; reference group and outgroup, and the outgroup in the analysis will use Yoruba or Mbuti as the outgroup according to different needs. If the result of the symmetrical f4 test yields a nonsignificant statistical value (absolute Z value < 3), it indicates that the count of derivative alleles (BABA) shared between the reference population and test population 1 is equal to the count of derived alleles (ABBA) shared between the

reference population and test population 2. Consequently, test population 1 and test population 2 form a genetically closer clade. A statistical significance positive value ($Z > 3$) indicates that the test population 1 contains relatively high genetic component related to the reference population. A statistical significance negative value ($Z < -3$) indicates that the test population 2 contains a relatively high genetic component associated with the reference population. The detection units of the affinity four-group test include reference group 1, reference group 2; test group and outgroup, and the statistically significant positive value indicates that the test group and reference group 1 share more alleles; whereas statistical significance negative value indicates that more alleles are shared between the test group and the reference group 2.

Multidimensional scaling

MDS analysis uses the similarity between pairs of samples to construct a low-dimensional space, so that the "distance" of the samples in this low-dimensional space and the similarity between samples in the high-dimensional space are as consistent as possible. MDS represents the physical spatial distributions of clusters in a population. This spatial distribution can represent the genetic differentiation pattern between populations to a certain extent. By directly comparing the two-dimensional vector map of MDS to analyze the relationship between genetic polymorphisms and the relationships between populations. The MDS analysis results are measured by the stress coefficient (stress). The MDS graph usually gives the stress value of the model, which is used to access whether the graph can accurately reflect the true distributions of data. It is generally believed that when stress < 0.2, it can be represented by a two-dimensional point graph of MDS, and its graph has certain explanatory significance; when stress < 0.1, it can be considered a good ranking; when stress < 0.05, it has a good representation. The MDS graph type is a scatter plot; the dots in the graph represent samples, and different colors/shapes represent group information of different samples. The distance between the sample points in the same group indicates the representation of the sample, and the distance between the grouped samples reflects the difference between groups. Currently, MDS analysis is commonly performed using softwares such as SPSS and R.

Principal component analysis

PCA is a widely used analysis method in forensic genetics. In 1995, Cavalli-Sforza et al. used PCA analysis for the first time to analyze complex population data on a global scale with multiple loci (Jolliffe, 2002). The study used dimensionality reduction method to preserve the variables between groups in the data as much as possible and intuitively reflect these variables on a two-dimensional plane. As a commonly used genetic statistics method, PCA calculates a new set of variables called principal components (PC) on the basis of the original data. Each new PC contributes a part of the variance and is arranged as PC1, PC2, … PCn (Cann, 1995). In the application of forensic genetics, PCA analysis

uses an intuitive visualization method to evaluate the degree of differentiation between populations and effectively reveals the genetic polymorphisms between populations through the relative positions of different populations. The research on European population structure stratification by Novembre and Lao showed that the PCA analysis method can effectively reveal the differentiations between populations (Pritchard et al., 2000). Although some researchers still believe that the purpose of the PCA analysis is to directly superimpose data variables on geographic space, leading to limited matching results that may not accurately reveal the relative degree of differentiation between populations. But in general, PCA analysis still provides an intuitive method to reveal the differentiation between populations in the field of forensic genetics. If there are large genetic differentiations between the study populations and the selected molecular genetic markers are sufficiently different, then when the PCA algorithm is used, the discrete clustering points formed by each sample in the two-dimensional space are enough to reflect the degree of genetic differentiation between populations. Common softwares for constructing PCA diagram are STRAF, MVSP, ORIGIN, R, PLINK, etc.

Phylogenetic tree

A phylogenetic tree is a genetic statistical method that uses a tree-like branch diagram to represent the genetic relationships between organisms or populations. In the field of forensic genetics, a phylogenetic tree is mainly constructed based on DNA sequences and DNA genotypes. The purpose of this algorithm is to reconstruct the ancestral sequence trait and measure the degree of divergence between different individuals or populations from the same ancestor using the branches of the dendrogram. The "leaf" node represents its degree of divergence. The branch of the dendrogram represents different subgroups, and the length of the branches represents the genetic distance between the populations. According to the presence or absence of "root", phylogenetic trees are divided into "rooted tree" and "unrooted tree". "Rooted trees" can also describe the temporal order of populations or genes, while "unrooted trees" mainly reflect the distances between populations (Fordyce et al., 2011; Seo et al., 2013). At present, the construction methods of phylogenetic trees include the following categories: distance method, maximum likelihood (ML), maximum parsimony (MP), and Bayesian inference (BI). The distance method calculates the distance matrix between the two sequences, which first repeatedly merges the two sequences with the shortest distance, and finally constructs the optimal tree. The algorithm is simple and easy to understand, and its calculation speed is fast. Unweighted pair-group method with arithmetic means (UPGMA) first classifies each sequence into a class, finds the two classes with the closest distance, and classifies them into a column, and then repeating this process until all clusters are added, and finally, a tree root is generated. However, this method is less commonly used nowadays. NJ is a commonly used method at present. It not only calculates the pairwise comparison distance but also minimizes the length of the entire

tree, thereby limiting the topology of the tree and overcoming the defect that the UPGMA algorithm requires a constant evolution rate. The NJ method is an algorithm based on the principle of minimal evolution (sequence homology). "Adjacent-joining" means that two taxa on a phylogenetic tree are connected by only one internal node. This method is relatively accurate, computationally fast, and makes minimal assumptions. However, because it treats all sites on the sequence indiscriminately and the evolutionary distances of the analyzed sequences are limited, the NJ method is suitable for short sequences with small evolutionary distances and fewer informative sites. The ML method is the most consistent algorithm with evolutionary facts, but it is computationally intensive and time-consuming, which is not conducive to large-scale population structure analysis. However, since its inference tree is not unique, it is not suitable for sequences with large variations, so the algorithm is only suitable for samples with long sequences of small differences in sequences and similar mutation rates. The MP method is established based on the hypothesis of the least number of base substitutions in the evolutionary process. It is originally used for the study of phenotypes. The principle of the MP method is to determine the shortest evolutionary tree by using the evolutionary principle with the least amount of change in traits among the most similar individuals. The BI method is based on the statistical inference of the evolutionary model to deal with the complex and close to the actual situation in the model. The algorithm integrates the existing phylogenetic data and intuitively reflects the reliability of each branch through the posterior probability. The advantage of the BI method is that it does not need to be tested by the bootstrap method, and is more sensitive to evolutionary model, which is suitable for large and complex datasets. The most commonly used softwares for phylogenetic trees are TreeMix software, MEGA software, and PHYLIP software.

Structure analysis

Structure analysis is a statistical method that uses genetic markers to infer the genetic structure of a population. The basic idea of STRUCTURE is to first obtain the genotypes of a series of samples. In most cases, researchers do not know how many subgroups the population actually contains. The number of subgroups of the population is the K value, and the preset number of subgroups is 1-n, that is, $K = 1 \sim n$. Then the software will use the Bayesian algorithm to calculate how the populations are grouped and what the ancestral composition of each individual is under the condition of simulating $K = 1 \sim n$.

At present, there are three softwares which can accomplish population structure analysis, namely STRUCTURE, ADMIXTURE and fast STRUCTURE. STRUCTURE software is developed based on the Bayesian algorithm and applied to genetic clustering statistics. STRUCTURE analysis is usually used to evaluate sample with mixed ancestral background and can provide a clearer composition ratio for individuals with mixed ancestral background. The software works based on the original genotyping results of STRs or SNPs and clusters the populations under the conditions of different population

clustering coefficients (*K*) (Reich et al., 2008; Fondevila et al., 2013). The STRUCTURE software can generate a cluster grouping coefficient matrix in one operation and compare the unknown sample with the reference population. This comparison does not consider the regional classification between the populations, but clusters the samples entirely based on genetic similarity. Therefore, this algorithm is often used to evaluate the discrimination efficiency of a specific combination of molecular genetic markers in related populations. If the cluster match the actual geographic source division, this set of molecular genetic markers can be considered to have good discriminating power for related populations. In addition, the STRUCTURE analysis is random, which means that multiple runs will produce different results (different values), but this does not affect the evaluation results of the genetic similarity between populations by this algorithm (Rosenberg, 2004; Jakobsson & Rosenberg, 2007). Using the DISTRUCT software (Kalinowski, 2011), the result of the STRUCTURE software can be visualized and presented in the form of an intuitive picture. In addition, the calculation result of STRUCTURE can be optimized by the CLIMPP software (Boeckmann et al., 2015), the stability of the population clustering coefficient *K* can be systematically measured by the online software HARVESTR, and the optimal *K* value corresponding to the population data can be screened out to avoid the presumption of the population affecting the results (Meng et al., 2011). In addition, in order to more comprehensively trace population evolution information, software such as Phase is usually used to perform haplotype phasing on the whole genome data to obtain the length or number of haplotype blocks required for subsequent analysis. Further, a comprehensive reconstruction analysis of population genetic affinity was performed using software such as fast STRUCTURE.

Another widely used software for genetic mixture model construction is ADMIXTOOLS, which has strong applicability to the large-throughput data generated by the NGS platform. ADMIXTOOLS is a software package developed in the laboratory of David Reich (Patterson et al., 2012). And it is designed for computing and testing population mixing statistics and validating hybridization events. Among them, qpWave and qpAdm are two programs on ADMIXTOOLS which can determine the minimum number of ancestral sources contributing to the studied population and calculate the proportions of ancestral components. Populations with distant kinships in the studied population and no recent admixtures among them are selected as an outgroup, and qpWave and qpAdm are used to construct a mixed model of the studied population. QpWave focuses more on exploring the use of the smallest ancestral population to interpret the genetic variation information contained in the target population. However, qpAdm mainly verifies the mixing ratio of a specific population under the formulation of the ancestral population. QpGraph is a hybrid modeling method that calculates allele frequency correlation patterns based on f2, f3, and f4 statistics simultaneously. The latest software version has been updated to the second-generation automatic or semiautomatic intelligent analysis system. Its genetic evolution model is superior to the traditionally used TreeMix.

With the development of NGS platform and the reduction of single-sample sequencing cost, the output of whole-genome data has currently increased exponentially.

MSMC, MSMC2, Momi, Momi2, and other softwares can analyze population segregation time, changes in effective population size, and complex population structure using genome-wide sample frequency maps or phased haplotype data. The above softwares are mainly developed based on the Linux system.

Learning practices of population study based on NGS platform for developing a 40 ancestry informative SNPs (AI-SNPs) multiplex panel for forensic ancestral inference on 17 different continental populations

Taking the construction of an ancestral prediction system as an example, we will briefly introduce some methods and basic workflow of population genetic study hereby. The general workflow of a population study is as follows: (1) Selecting a set of SNP loci associated with different geographical origins within the scope of the whole human genome; (2) Analyzing the genetic structures of 17 continental populations from AFR, EUR, and EAS based on the raw genotype data of the selected SNPs provided by the 1000 Genome Project database; (3) Evaluating the efficacy of the selected same loci when predicting the biogeographical origins of continents, as well as investigating the genetic relationships and backgrounds of the 17 populations from the three continents.

Assessing the contributions of the 40 AI-SNPs for ancestral information inference

The criteria for the selection of SNP loci are as follows: (1) The allele frequency differences (delta, δ) between any pairwise continental populations from AFR, EUR, and EAS are greater than 0.4; (2) All of the selected loci need to be not only biallele SNP loci but also need to be located in the intron region, so as to reduce the risks of bias in genotyping caused by diseases, etc.; (3) The physical distances of pairwise loci from the same chromosome must be greater than 10 Mb, and there is no linkage between the selected loci; (4) All of the SNPs must not display the deviations from Hardy-Weinberg equilibrium in the East Asian populations; (5) No linkage disequilibrium is found between any two SNP loci.

Based on the above criteria, 40 AI-SNP loci for ancestral inference of populations from the AFR, EUR, and EAS are finally confirmed. Utilizing the 1000 Genome Project data of the 17 reference populations from three continents, the selected 40 AI-SNPs are evaluated for their efficacies in ancestral inference. The Cos^2 values of 40 AI-SNPs obtained from the PCA analysis based on the genotype frequency data of the studied populations are shown in Fig. 8.1. The Cos^2 values are commonly used to compare the contribution of each SNP to analyze the population differentiations, and the larger the Cos^2 value is, the farther the corresponding SNP is from the origin point and closer to the periphery, indicating that this SNP performs better in explaining the population differentiation. As shown in this figure, the Cos^2 values range from 0.94 to 0.99, and all of the selected 40 AI-SNPs roughly distribute around the periphery of the circle, indicating

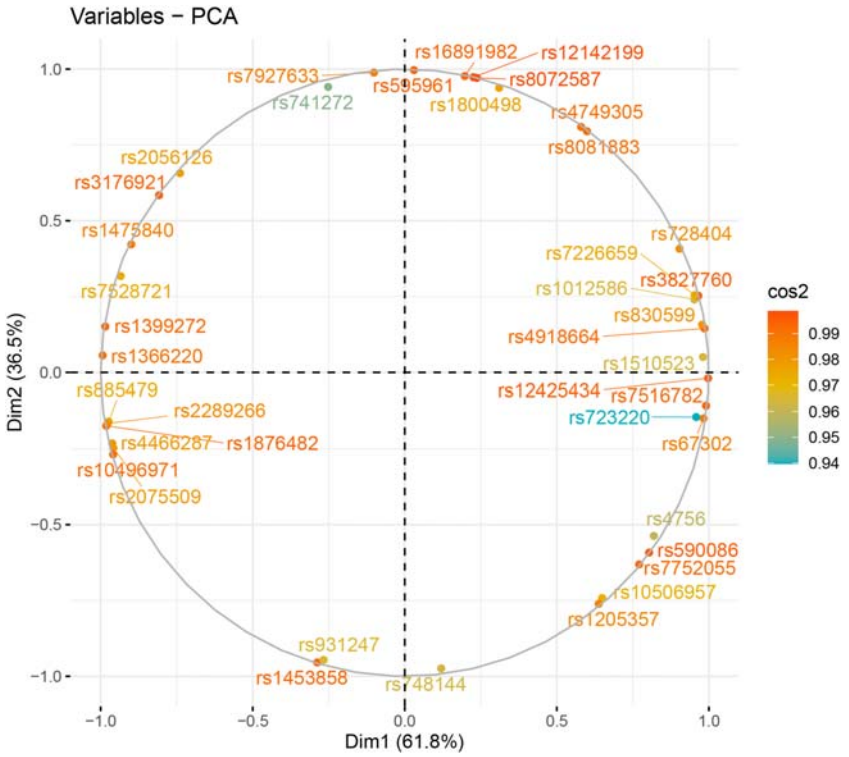


Figure 8.1 Cos^2 values of 40 AI-SNPs in 17 different continental populations.

that the contribution levels of these 40 AI-SNPs are relatively high and balanced. This indicates that their combination effect is good, and further enhances their applicability to ancestry-informative inference. This balanced distribution of contributions from the 40 AI-SNPs ensures that the prediction is not dominated by a single SNP, making the overall inference performance more robust and accurate.

Gene frequency statistics of 40 AI-SNP loci in the 17 populations from three continents

The gene frequencies of selected 40 AI-SNPs in the 17 continental populations from AFR, EUR, and EAS are statistically analyzed, and the related frequency heat map is illustrated in Fig. 8.2. The color of each grid represents the allelic frequency value of the corresponding SNP in the assigned population, and according to the yellow-blue gradient scale on the right, the yellower the grid, the higher the frequency, and vice versa. The color blocks on the left indicate the continental origins of these populations, with African, European, and East Asian populations labeled as green, blue, and pink, respectively. As shown in the figure, the frequency distributions of the selected 40 AI-SNPs display significant differences among the three continental populations. It can be seen from the frequency-based clustering results on the left edge that the 17 studied

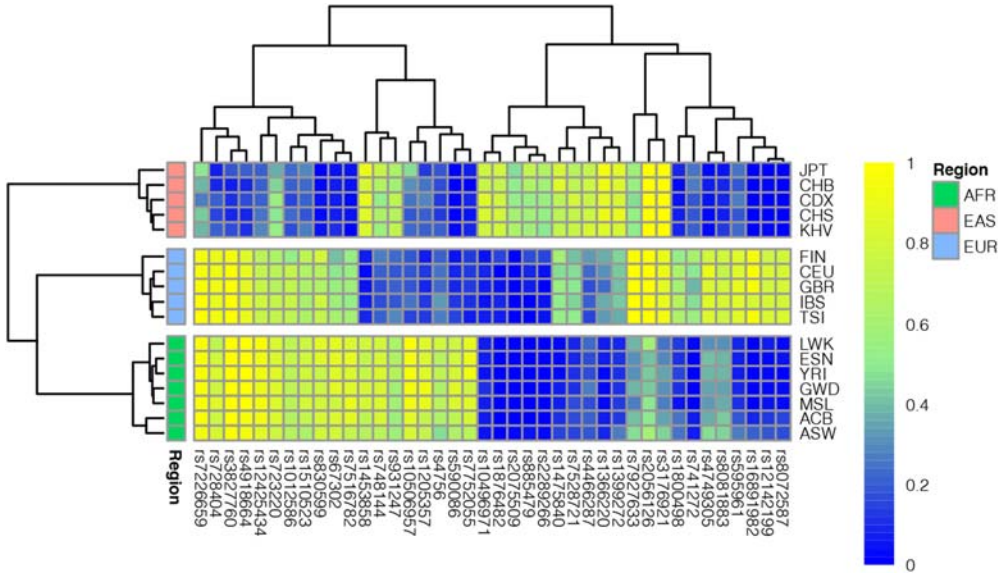


Figure 8.2 Gene frequency heatmap of 40 AI-SNPs in 17 different continental populations. The population names and the corresponding abbreviations are as follows: *ACB*, African Caribbean in Barbados; *ASW*, African Ancestry in Southwest US; *CDX*, Xishuangbanna Dai Chinese; *CEU*, Utah residents with Northern and Western European ancestry (France); *CHB*, Beijing Han Chinese; *CHS*, Southern Han Chinese; *ESN*, Esan in Nigeria; *FIN*, Finnish in Finland; *GBR*, British in England and Scotland; *GWD*, Gambian in Western Division-Mandinka; *IBS,* Iberian populations in Spain; *JPT*, Tokyo Japanese; *KHV*, Ho Chi Minh Kinh; *LWK*, Luhya in Webuye, Kenya; *MSL*, Mende in Sierra Leone; *TSI*, Toscani in Italy; *YRI*, Yoruba in Ibadan, Nigeria.

populations are mainly clustered into three sub-branches, and the populations from the same continent are obviously clustered together and separated from other intercontinental populations. Therefore, the selected 40 AI-SNPs may have the potential ability in distinguishing African, East Asian, and European populations.

Investigation of the genetic relationships among the 17 populations from three continents using 40 AI-SNPs

Population differentiation measures of the 17 continental populations

The pairwise F_{ST} is a basic measure of population differentiation, and the related pairwise F_{ST} heat map with two symmetrical clustering trees of the 17 studied populations is demonstrated in Fig. 8.3A. The two symmetrical clustering trees of F_{ST} are both divided into the same three sub-branches, with the East Asian populations cluster located at the top, the European populations cluster in the middle, and the African populations cluster at the bottom. The populations from the same continent cluster with each other and are separated from the populations on other continents. The pairwise F_{ST} values between the 17 continental populations range from 0 to 0.911, with an average of 0.508. The smallest F_{ST} value is found between the ESN and YRI populations from AFR, whereas the largest F_{ST} value is observed between the ESN population from AFR and the CHS from EAS. In African populations,

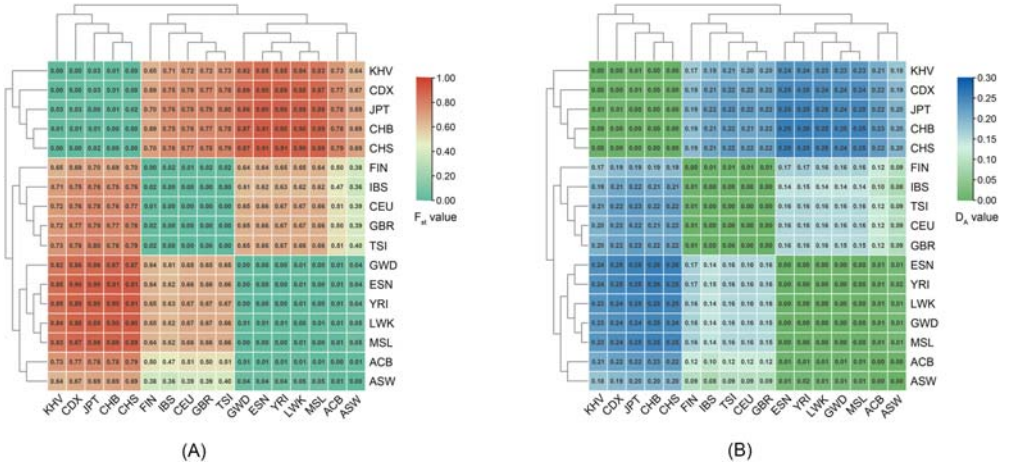


Figure 8.3 Heatmaps of the pairwise F_{ST} (A) and D_A (B) values in 17 different continental populations calculated with the 40 AI-SNPs genotypes.

the F_{ST} values range from 0 to 0.048, with an average of 0.017. In European populations, the F_{ST} values vary from 0.001 to 0.018, with a mean of 0.008. As for East Asian populations, the F_{ST} values range from 0.002 to 0.030, with an average of 0.012. In summary, the results of F_{ST} analysis reveal less genetic differentiations and closer genetic relationships among populations from the same continent, while greater differentiation coefficients and more distant genetic relationships were observed among populations from different continents.

As another widely used measure of population genetic distance is based on the infinite gene mutation model, the pairwise D_A value matrix of 17 populations is also presented as a heatmap (Fig. 8.3B). According to the pairwise D_A values, the populations from the three continents cluster into three sub-branches, and the populations with the same geographical origin cluster together and separate from the populations on other continents. The pairwise D_A values of the 17 populations range from 0.0015 to 0.2594, with the highest value found between the ESN from AFR and the CHB from EAS, and the lowest value observed between the two European populations FIN and CEU. Similar to the F_{ST} analysis results, the distribution patterns of lower D_A values among populations from the same continent indicate closer genetic distances and possibly more gene exchanges, while the patterns of higher D_A values among populations from different continents suggest longer genetic distances among these populations than among those populations within the same continent.

PCA analyses of the 17 continental populations based on 40 AI-SNPs

PCA analyses of the studied 17 populations from three continents are performed at the population level and individual level, respectively (Fig. 8.4). As shown in the PCA analysis of population level, the first two PCs account for 95.1% of the total variation (61.76% PC1 and 36.48% PC2). Points with red, blue, and green colors represent the populations from

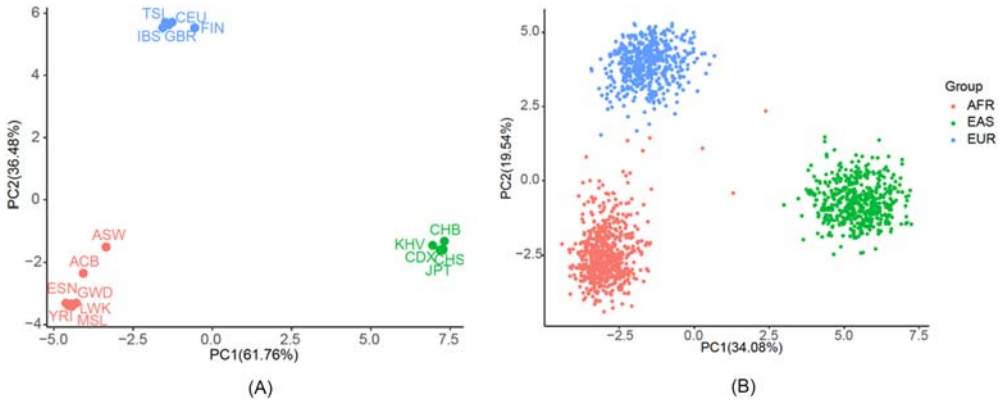


Figure 8.4 Population (A) and individual (B) levels of PCA analyses in the 17 continental populations using 40 AI-SNPs.

AFR, EUR, and EAS, respectively, and it can be seen from Fig. 8.4A that the populations from the same geographical region cluster together. AFR and EAS populations group in the left and right quadrants, respectively, while the EUR populations cluster in the middle. In Fig. 8.4B, 1668 individuals are classified according to their continental origins, which are labeled with different colors; in other words, red, blue, and green represent individuals from AFR, EUR, and EAS, respectively. Individuals from the same continent cluster together and separate from populations with different continental origins.

Phylogenetic analyses of the 17 continental populations based on 40 AI-SNPs

Phylogenetic analysis is an important method of genetic studies, which applies the branch diagram to illustrate the genetic relationships between different species. In this section, MEGA software and TreeMix software are used to reconstruct the phylogenetic trees of the studied populations on the basis of NJ and ML algorithms and to further analyze the phylogenetic relationships among the 17 populations.

In the NJ tree based on pairwise F_{ST} values of 17 populations, the populations from the same continent are annotated by the same background color, and these populations mainly distribute in three major branches of the tree, namely the African, European, and East Asian branches (Fig. 8.5A). In each major branch, the subgroups from the same continent cluster together. As for the unrooted ML tree based on ML ratio (Fig. 8.5B), populations from the same continent are also labeled with the same color according to the caption on the right. It can be inferred from the figure that the 7 populations from AFR cluster together at the top of the tree; the 5 populations clustered in the middle are from EUR; while the East Asian populations cluster at the bottom of the tree. Overall, both of the phylogenetic trees constructed based on NJ and ML algorithms can effectively distinguish the populations from the three continents, but there are some inconsistencies when it comes to the more in-depth estimations of relationships within populations from the same continent.

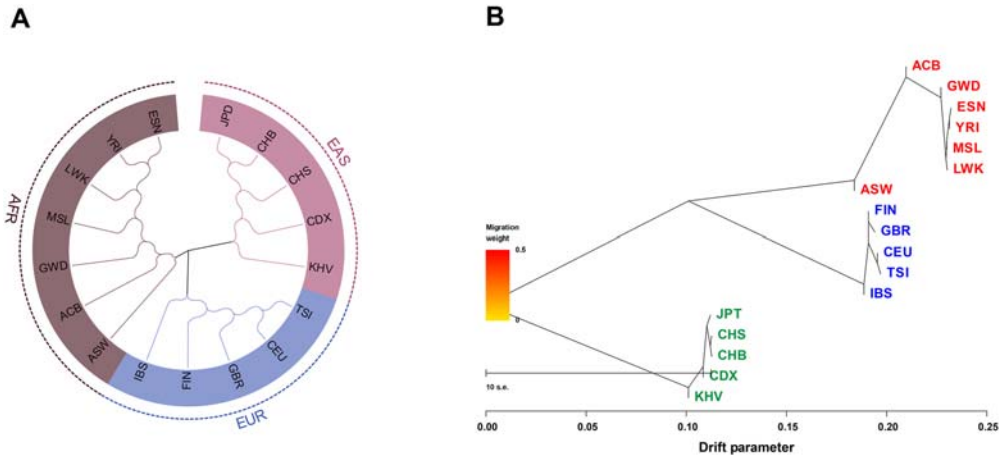


Figure 8.5 The NJ tree (A) and inferred ML phylogenetic tree (B) were conducted based on 17 continental populations.

In the NJ tree, ESN and YRI show a close genetic distance among the 7 populations in AFR and then cluster with LWK, MSL, GWD, ACB, and ASW successively. In the ML tree, the African populations are divided into three branches, wherein the ESN, YRI, MSL, and LWK display closer genetic distances and then cluster together with the GWD, ACB, and ASW, respectively. For populations from EUR and EAS, the genetic relationships among different populations estimated by the NJ and ML methods are similar. Within the European populations, the closest genetic distance is observed between TSI and CEU, and the two populations then cluster with GBR, FIN, and IBS into the European branch of the NJ tree. As for the ML tree, TSI still has the closest genetic relationship with CEU, forming a small subbranch, while GBR and FIN display a close genetic distance, forming another small subbranch. Then, these two sub-branches cluster together with IBS into the European branch. Among the five populations from EAS, JPT and CHB have the closest genetic distance in the NJ tree and subsequently cluster together with CHS, CDX, and KHV, forming the East Asian branch. As for the ML tree, Chinese Han populations of CHS and CHB share the most recent ancestry and later cluster together with JPT, CDX, and KHV.

Structure analysis of the 17 continental populations with 40 AI-SNPs

Based on 40 AI-SNPs, a STRUCTURE analysis is performed in the 17 continental populations (Fig. 8.6). According to the highest estimated values of Delta K and Prob (K), the optimal K is eventually determined to be 3. When $K = 2$, populations with the dominant East Asian ancestral component (green) are separated from the others with different geographical origins, while the African (orange) and European (red) ancestral components are later identified when $K = 3$. However, no new groups with novel ancestral component patterns are further determined by the increasing numbers of K (4—7). As

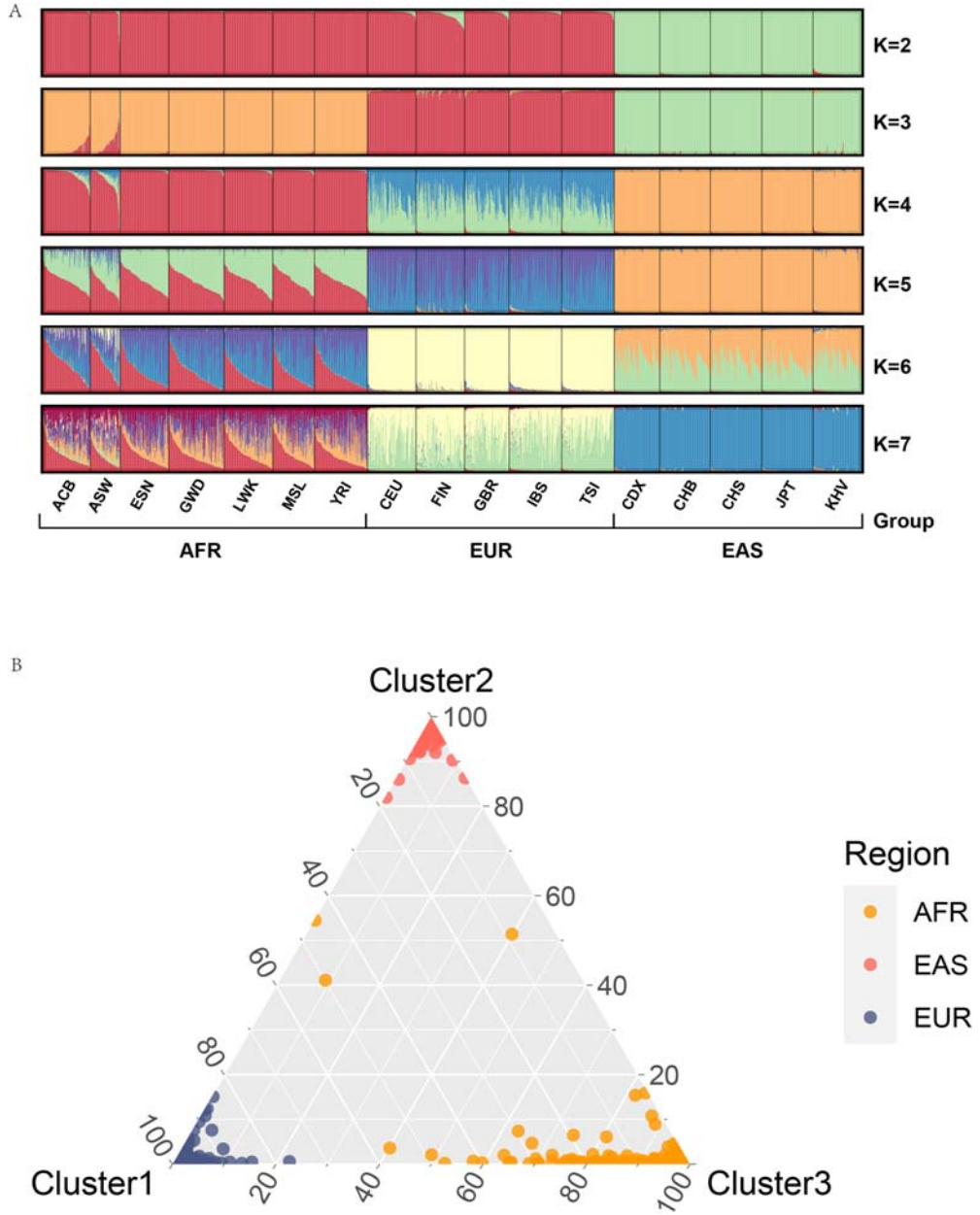


Figure 8.6 STRUCTURE analyses ($K = 2—7$) on individual level based on 40 AI-SNPs genotype data of the 1668 unrelated individuals from 17 continental populations (A); Cluster analysis for the results of STRUCTURE analysis on individual level when $K = 3$ (B).

shown in the cluster analysis (Fig. 8.6), individuals from East Asian (red), European (blue), and African (yellow) populations distribute at the three corners of the ternary diagram. Individuals from the same continent congregate with each other and separate from individuals with other continental origins. STRUCTURE and cluster results further demonstrate that the selected 40 AI-SNPs have a promising performance in distinguishing the African, European, and East Asian populations.

Ancestral inference using the selected 40 AI-SNPs

Based on the genotyping data of 17 populations from the three continents, a supervised learning algorithm of "Random Forest" is applied to evaluate the efficacy of the combination of 40 AI-SNPs system as a classifier of continental populations by randomly selected 75% of all the individuals from each continent as the training set and the remaining 25% as the testing set. As a result, the selected 40 AI-SNPs are almost 100% accurate (95% CI: 0.9912−1) when distinguishing the studied populations from the three continents. The prediction accuracy of the random forest model based on 40 AI-SNPs for populations from the three continents is shown in Table 8.6. Except for one African individual who is incorrectly assigned to be European, the geographical origins of the remaining 415 individuals from the testing set are correctly identified. Evaluation metrics of the random forest predictive model of ancestral origin inference based on the 40 AI-SNPs are displayed in Table 8.7. It can be concluded from the table that the constructed model exhibits considerably high sensitivity and specificity when predicting the testing set, and the false negative and false positive rates are rather low.

Table 8.6 Predictions of ancestral origins for populations from the three continents based on the random forest model using 40 AI-SNPs.

Reference prediction	Africa	East Asia	Europe
Africa	164	0	0
East Asia	0	126	0
Europe	1	0	125

Table 8.7 Evaluation metrics of the random forest predictive model of ancestral origins based on the 40 AI-SNPs.

Class	African	East Asian	European
Sensitivity	1	1	1
Specificity	1	1	1
Positive prediction value	1	1	1
Negative prediction value	1	1	1
Prevalence	0.3966	0.3029	0.3005
Detection rate	0.3966	0.3029	0.3005
Detection prevalence	0.3966	0.3029	0.3005
Balanced accuracy	1	1	1

Summary

1. Forensic population genetic study aims to investigate the variations of gene frequencies and genotype frequencies in different populations by mathematical and statistical methods; to study the genetic differentiations within and between populations; to analyze the genetic structures of populations; to survey their genetic backgrounds; and eventually to lay the foundation for forensic applications.
2. Before each genetic marker or multiplex panel is applied to forensic practice, it is necessary to systematically evaluate its genetic polymorphism and population applicability, which aims to develop reasonable analytical method for forensic application based on the result of population study.
3. As the currently mainstream molecular markers for population study on the NGS platform, STRs, SNPs, mtDNA, MHs, and compound markers (i.e., SNP-STR, InDel-SNP, InDel-STR) can play an important role in the applications of forensic individual identification, kinship analysis, ancestral information inference, and so on.
4. Estimation of F_{ST}, F test, MDS analysis, PCA, phylogenetic tree reconstruction, and genetic admixture predictive model are the main analytical approaches for forensic population genetic research.
5. Based on the NGS platform, a combination of 40 AI-SNPs is studied, which can be used for forensic ancestral information inference of the studied 17 populations from the three continents of EUR, AFR, and EAS.

Acknowledgments

This work has received fundings from the Major Program of the National Natural Science Foundation of China (No. 82293652 and No. 82293650) and the Key Program of the National Natural Science Foundation of China (Grant numbers: 81930055).

References

Aalbers, S. E., Hipp, M. J., Kennedy, S. R., & Weir, B. S. (2020). Analyzing population structure for forensic STR markers in next generation sequencing data. *Forensic Science International Genetics, 49*, 102364. https://doi.org/10.1016/j.fsigen.2020.102364. Epub 2020 Aug 12. PMID: 32805606; PMCID: PMC7655679.

Al-Asfi, M., McNevin, D., Mehta, B., Power, D., Gahan, M. E., & Daniel, R. (2018). Assessment of the Precision ID Ancestry panel. *International Journal of Legal Medicine, 132*(6), 1581−1594. https://doi.org/10.1007/s00414-018-1785-9. Epub 2018 Mar 19. PMID: 29556719.

Ambers, A., Bus, M. M., King, J. L., Jones, B., Durst, J., Bruseth, J. E., Gill-King, H., & Budowle, B. (2020). *Forensic Science International, 306*. https://doi.org/10.1016/j.forsciint.2019.110050

Avila, E., Felkl, A. B., Graebin, P., Nunes, C. P., & Alho, C. S. (2019). *Forensic Science International: Genetics, 40*, 74−84. https://doi.org/10.1016/j.fsigen.2019.02.012

Avila, E., Graebin, P., Chemale, G., Freitas, J., Kahmann, A., & Alho, C. S. (2019). Full mtDNA genome sequencing of Brazilian admixed populations: A forensic-focused evaluation of a MPS application as an alternative to Sanger sequencing methods. *Forensic Science International, 42*, 154−164. https://doi.org/10.1016/j.fsigen.2019.07.004. Epub 2019 Jul 9. PMID: 31325893.

Avila, E., Speransa, P. A., Lindholz, C. G., Kahmann, A., & Alho, C. S. (2022). *Forensic Science International: Genetics, 57*, 102650. https://doi.org/10.1016/j.fsigen.2021.102650

Barrio, P. A., Martín, P., Alonso, A., Müller, P., Bodner, M., Berger, B., Parson, W., & Budowle, B. (2019). *Forensic Science International: Genetics, 42*, 49–55. https://doi.org/10.1016/j.fsigen.2019.06.009

Boeckmann, B., Marcethouben, M., Rees, J. A., Forslund, K., Huerta-Cepas, J., Muffato, M., Yilmaz, P., Xenarios, I., Bork, P., Lewis, S., Gabaldón, T., & Quest for Orthologs Species Tree Working Group. (2015). Quest for orthologs entails quest for tree of life: In search of the gene stream. *Genome Biology & Evolution, 7*(7), 1988–1999. https://doi.org/10.1093/gbe/evv121

Bulbul, O., & Filoglu, G. (2018). *Electrophoresis, 39*(21), 2743–2751. https://doi.org/10.1002/elps.201800243

Bulbul, O., Speed, W. C., Gurkan, C., Soundararajan, U., Rajeevan, H., Pakstis, A. J., & Kidd, K. K. (2018). *Forensic Science International: Genetics, 35*, 14–20. https://doi.org/10.1016/j.fsigen.2018.03.010

Cann, R. L. (1995). The history and geography of human ggenes. *The American Journal of Human Genetics, 56*(1).

Chaitanya, L., Ralf, A., van Oven, M., Kupiec, T., Chang, J., Lagacé, R., & Kayser, M. (2015). *Human Mutation, 36*(12), 1236–1247. https://doi.org/10.1002/humu.22905

Chen, P., Yin, C., Li, Z., Pu, Y., Yu, Y., Zhao, P., Chen, D., Liang, W., Zhang, L., & Chen, F. (2018). *Forensic Science International: Genetics, 35*, 149–155. https://doi.org/10.1016/j.fsigen.2018.05.003

Chen, P., Zhu, W., Tong, F., Pu, Y., Yu, Y., Huang, S., Li, Z., Zhang, L., Liang, W., & Chen, F. (2019). *International Journal of Legal Medicine, 133*(4), 983–988. https://doi.org/10.1007/s00414-018-1881-x

Chen, Q.-F., Kang, K.-L., Song, J.-J., Zhang, C., Yu, Z.-L., Zhao, G.-B., Wu, H., Ji, A.-Q., Ye, J., & Wang, L. (2021). *Forensic Science International: Genetics, 51*, 102424. https://doi.org/10.1016/j.fsigen.2020.102424

Chen, P., Deng, C., Li, Z., Pu, Y., Yang, J., Yu, Y., Li, K., Li, D., Liang, W., Zhang, L., & Chen, F. (2019). A microhaplotypes panel for massively parallel sequencing analysis of DNA mixtures. *Forensic Science International Genetics*, *40*, 140–149. https://doi.org/10.1016/j.fsigen.2019.02.018. Epub 2019 Feb 22. PMID: 30836262.

Chen, C., Jin, X., Zhang, X., Zhang, W., Guo, Y., Tao, R., Chen, A., Xu, Q., Li, M., Yang, Y., & Zhu, B. (2021). Comprehensive insights into forensic features and genetic background of Chinese Northwest Hui Group using six distinct categories of 231 molecular markers. *Frontiers in Genetics*, *12*, 705753. https://doi.org/10.3389/fgene.2021.705753. PMID: 34721519; PMCID: PMC8555763.

Cooley, A. M., Meiklejohn, K. A., Damaso, N., Robertson, J. M., & Dawson Cruz, T. (2021). *SLAS Technology, 26*(1), 103–112. https://doi.org/10.1177/2472630320954180

Dai, W., Pan, Y., Sun, X., Wu, R., Li, L., & Yang, D. (2019). *Scientific Reports, 9*(1). https://doi.org/10.1038/s41598-019-55282-9

Dash, H. R., Kaitholia, K., Kumawat, R. K., Singh, A. K., Shrivastava, P., Chaubey, G., & Das, S. (2021). *Scientific Reports, 11*(1). https://doi.org/10.1038/s41598-021-02690-5

Dash, H. R., Avila, E., Jena, S. R., Kaitholia, K., Agarwal, R., Alho, C. S., Srivastava, A., & Singh, A. K. (2022). *International Journal of Legal Medicine, 136*(2), 465–473. https://doi.org/10.1007/s00414-021-02742-5

de la Puente, M., Phillips, C., Santos, C., Fondevila, M., Carracedo, & Lareu, M. V. (2017). *Forensic Science International: Genetics, 28*, 35–43. https://doi.org/10.1016/j.fsigen.2017.01.012

de la Puente, M., Phillips, C., Xavier, C., Amigo, J., Carracedo, A., Parson, W., & Lareu, M. V. (2020). Building a custom large-scale panel of novel microhaplotypes for forensic identification using MiSeq and Ion S5 massively parallel sequencing systems. *Forensic Science International Genetics*, *45*, 102213. https://doi.org/10.1016/j.fsigen.2019.102213. Epub 2019 Nov 30. PMID: 31835179.

de la Puente, M., Ruiz-Ramírez, J., Ambroa-Conde, A., Xavier, C., Pardo-Seco, J., Álvarez-Dios, J., Freire-Aradas, A., Mosquera-Miguel, A., Gross, T. E., Cheung, E. Y. Y., Branicki, W., Nothnagel, M., Parson, W., Schneider, P. M., Kayser, M., Carracedo, Á., Lareu, M. V., & Phillips, C. (2021). Development and evaluation of the ancestry informative marker panel of the VISAGE basic tool. *Genes (Basel), 12*, 1284. https://doi.org/10.3390/genes12081284

Delest, A., Godfrin, D., Chantrel, Y., Ulus, A., Vannier, J., Faivre, M., Hollard, C., & Laurent, F.-X. (2020). *Forensic Science International: Genetics, 47*, 102304. https://doi.org/10.1016/j.fsigen.2020.102304

Devesse, L., Davenport, L., Borsuk, L., Gettings, K., Mason-Buck, G., Vallone, P. M., Syndercombe Court, D., & Ballard, D. (2020). *Forensic Science International: Genetics, 48*, 102356. https://doi.org/10.1016/j.fsigen.2020.102356

Eduardoff, M., Xavier, C., Strobl, C., Casas-Vargas, A., & Parson, W. (2017). *Genes, 8*(10). https://doi.org/10.3390/genes8100237

Fan, H., Du, Z., Wang, F., Wang, X., Wen, S. Q., Wang, L., Du, P., Liu, H., Cao, S., Luo, Z., Han, B., Huang, P., Zhu, B., & Qiu, P. (2021). *International Journal of Legal Medicine, 135*(4), 1295–1317. https://doi.org/10.1007/s00414-021-02590-3

Fan, H., Wang, L., Liu, C., Lu, X., Xu, X., Ru, K., Qiu, P., Liu, C., & Wen, S. Q. (2022). Development and validation of a novel 133-plex forensic STR panel (52 STRs and 81 Y-STRs) using single-end 400 bp massive parallel sequencing. *International Journal of Legal Medicine*, *136*(2), 447-464. https://doi.org/10.1007/s00414-021-02738-1. Epub 2021 Nov 6. PMID: 34741666.

Fondevila, M., Phillips, C., Santos, C., Aradas, A. F., Vallone, P. M., Butler, J. B., & Carracedo, A. (2013). Revision of the SNPforID 34-plex forensic ancestry test: Assay enhancements, standard reference sample genotypes and extended population studies. *Forensic Science International Genetics,* 7(1), 63–74. https://doi.org/10.1016/j.fsigen.2012.06.007

Fordyce, S. L., Ávila-Arcos, M. C., Rockenbauer, E., Børsting, C., Frank-Hansen, R., Petersen, F. T., Willerslev, E., Hansen, A. J., Morling, N., & Gilbert, T. P. (2011). High-throughput sequencing of core STR loci for forensic genetic investigations using the Roche Genome Sequencer FLX platform. *BioTechniques, 51*(2), 127–133. https://doi.org/10.2144/000113721

Gandotra, N., Speed, W. C., Qin, W., Tang, Y., Pakstis, A. J., Kidd, K. K., & Scharfe, C. (2020). *Forensic Science International: Genetics, 47*, 102275. https://doi.org/10.1016/j.fsigen.2020.102275

Gettings, K. B., Kiesler, K. M., Faith, S. A., Montano, E., Baker, C. H., Young, B. A., Guerrieri, R. A., & Vallone, P. M. (2016). *Forensic Science International: Genetics, 21*, 15–21. https://doi.org/10.1016/j.fsigen.2015.11.005

Gorden, E. M., Sturk-Andreaggi, K., & Marshall, C. (2018). *Forensic Science International: Genetics, 34*, 257–264. https://doi.org/10.1016/j.fsigen.2018.02.015

Guo, F., Zhou, Y., Song, H., Zhao, J., Shen, H., Zhao, B., Liu, F., & Jiang, X. (2016). *Forensic Science International: Genetics, 25*, 73–84. https://doi.org/10.1016/j.fsigen.2016.07.021

Guo, Y. X., Jin, X. Y., Xia, Z. Y., Chen, C., Cui, W., & Zhu, B. F. (2020). *Electrophoresis, 41*(9), 649–656. https://doi.org/10.1002/elps.201900231

Gusmão, L., Butler, J. M., Linacre, A., Parson, W., Roewer, L., Schneider, P. M., & Carracedo, A. (2017). Revised guidelines for the publication of genetic population data. *Forensic Science International Genetics*, *30*, 160–163. https://doi.org/10.1016/j.fsigen.2017.06.007. Epub 2017 Jun 23. PMID: 28673498.

He, G., Liu, J., Wang, M., Zou, X., Ming, T., Zhu, S., Yeh, H. Y., Wang, C., Wang, Z., & Hou, Y. (2021). Massively parallel sequencing of 165 ancestry-informative SNPs and forensic biogeographical ancestry inference in three southern Chinese Sinitic/Tai-Kadai populations. *Forensic Science International Genetics*, *52*, 102475. https://doi.org/10.1016/j.fsigen.2021.102475. Epub 2021 Feb 2. PMID: 33561661.

Holland, M. M., Makova, K. D., & McElhoe, J. A. (2018). Deep-coverage MPS analysis of heteroplasmic variants within the mtGenome allows for frequent differentiation of maternal relatives. *Genes (Basel)*, *9*(3), 124. https://doi.org/10.3390/genes9030124. PMID: 29495418; PMCID: PMC5867845.

Houston, R., Mayes, C., King, J. L., Hughes-Stamm, S., & Gangitano, D. (2018). *Electrophoresis, 39*(22), 2906–2911. https://doi.org/10.1002/elps.201800152

Hussing, C., Bytyci, R., Huber, C., Morling, N., & Børsting, C. (2019). *International Journal of Legal Medicine, 133*(2), 325–334. https://doi.org/10.1007/s00414-018-1854-0

Huszar, T. I., Jobling, M. A., & Wetton, J. H. (2018). *Forensic Science International: Genetics, 35*, 97–106. https://doi.org/10.1016/j.fsigen.2018.03.012

Jakobsson, M., & Rosenberg, N. A. (2007). CLUMPP: A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics, 23*(14), 1801–1806.

Jin, X. Y., Cui, W., Chen, C., Guo, Y. X., Zhang, X. R., Xing, G. H., Lan, J. W., & Zhu, B. F. (2020). *Electrophoresis, 41*(13–14), 1230–1237. https://doi.org/10.1002/elps.201900451

Jin, X., Zhang, X., Shen, C., Liu, Y., Cui, W., Chen, C., Guo, Y., & Zhu, B. (2020). *Genes, 11*(9), 1–14. https://doi.org/10.3390/genes11091027

Jin, X. Y., Liu, Y. F., Cui, W., Chen, C., Zhang, X. R., Huang, J., & Zhu, B. F. (2022). Development a multiplex panel of AISNPs, multi-allelic InDels, microhaplotypes, and Y-SNP/InDel loci for multiple forensic purposes via the NGS. *Electrophoresis*, *43*(4), 632–644. https://doi.org/10.1002/elps.202100253. Epub 2022 Jan 6. PMID: 34859475.

Jolliffe, I. T. (2002). Principal component analysis. *Journal of Marketing Research, 87*(100), 513.
Kalinowski, S. T. (2011). The computer program STRUCTURE does not reliably identify the main genetic clusters within species: Simulations and implications for human population structure. *Heredity, 106*(4), 625–632. https://doi.org/10.1038/hdy.2010.95
Khubrani, Y. M., Hallast, P., Jobling, M. A., & Wetton, J. H. (2019). *Forensic Science International: Genetics, 43*, 102164. https://doi.org/10.1016/j.fsigen.2019.102164
Kidd, K. K., Bulbul, O., Gurkan, C., Dogan, M., Dogan, S., Neophytou, P. I., Cherni, L., Gurwitz, D., Speed, W. C., Murtha, M., Kidd, J. R., & Pakstis, A. J. (2021). *Forensic Science International: Genetics, 53*, 102528. https://doi.org/10.1016/j.fsigen.2021.102528
Kim, E. H., Lee, H. Y., Kwon, S. Y., Lee, E. Y., Yang, W. I., & Shin, K. J. (2017). *Forensic Science International: Genetics, 30*, 134–140. https://doi.org/10.1016/j.fsigen.2017.07.001
Kim, S. Y., Lee, H. C., Chung, U., Ham, S. K., Lee, H. Y., Park, S. J., Roh, Y. J., & Lee, S. H. (2018). *Electrophoresis, 39*(21), 2702–2707. https://doi.org/10.1002/elps.201800090
Kim, M. Y., Cho, S., Lee, J. H., Seo, H. J., & Lee, S. D. (2018). Detection of innate and artificial mitochondrial DNA heteroplasmy by massively parallel sequencing: Considerations for analysis. *Journal of Korean Medical Science*, *33*(52), e337. https://doi.org/10.3346/jkms.2018.33.e337. PMID: 30584415; PMCID: PMC6300661.
Kimura, M. (1968). Evolutionary rate at the molecular level. *Nature, 217*, 624–626. https://doi.org/10.1038/217624a0.
King, J. L., Churchill, J. D., Novroski, N. M. M, Zeng, X., Warshauer, D. H., Seah, L. H., & Budowle, B. (2018). Increasing the discrimination power of ancestry- and identity-informative SNP loci within the ForenSeq DNA Signature Prep Kit. *Forensic Science International Genetics*, *36*, 60–76. https://doi.org/10.1016/j.fsigen.2018.06.005. Epub 2018 Jun 6. PMID: 29935396.
Kureshi, A., Li, J., Wen, D., Sun, S., Yang, Z., & Zha, L. (2020). *Royal Society Open Science, 7*(5), 191937. https://doi.org/10.1098/rsos.191937
Kwon, S. Y., Lee, H. Y., Kim, E. H., Lee, E. Y., & Shin, K. J. (2016). *Forensic Science International: Genetics, 25*, 132–141. https://doi.org/10.1016/j.fsigen.2016.08.010
Lee, J. H., Cho, S., Kim, M. Y., Shin, D. H., Rakha, A., Shinde, V., & Lee, S. D. (2018). *Legal Medicine, 34*, 41–47. https://doi.org/10.1016/j.legalmed.2018.08.007
Lee, J. H., Kim, J., Kim, H., Kim, H. S., & Kim, E. (2021). *Genes and Genomics, 43*(2), 133–140. https://doi.org/10.1007/s13258-020-01033-4
Li, L., Wang, Y., Yang, S., Xia, M., Yang, Y., Wang, J., Lu, D., Pan, X., Ma, T., Jiang, P., Yu, G., Zhao, Z., Ping, Y., Zhou, H., Zhao, X., Sun, H., Liu, B., Jia, D., Li, C., … Li, S. (2017). *Forensic Science International: Genetics, 28*, 118–127. https://doi.org/10.1016/j.fsigen.2017.02.005
Li, R., Zhang, C., Li, H., Wu, R., Li, H., Tang, Z., Zhen, C., Ge, J., Peng, D., Wang, Y., Chen, H., & Sun, H. (2018). *International Journal of Legal Medicine, 132*(4), 997–1006. https://doi.org/10.1007/s00414-017-1706-3
Li, Ran, Wu, R., Li, H., Zhang, Y., Peng, D., Wang, N., Shen, X., Wang, Z., & Sun, H. (2020). *Forensic Science International: Genetics, 45*, 102225. https://doi.org/10.1016/j.fsigen.2019.102225
Li, H., Zhang, C., Song, G., Ma, K., Cao, Y., Zhao, X., Yang, Q., & Xie, J. (2021). *Molecular Genetics and Genomic Medicine, 9*(4). https://doi.org/10.1002/mgg3.1626
Liu, J., Wang, Z., He, G., Zhao, X., Wang, M., Luo, T., Li, C., & Hou, Y. (2018). *Forensic Science International: Genetics, 35*, 141–148. https://doi.org/10.1016/j.fsigen.2018.05.002
Liu, Q., Ma, G., Du, Q., Lu, C., Fu, L., Wang, Q., Fu, G., Li, S., & Cong, B. (2020). *International Journal of Legal Medicine, 134*(6), 2005–2014. https://doi.org/10.1007/s00414-020-02295-z
Lopopolo, M., Børsting, C., Pereira, V., & Morling, N. (2016). *American Journal of Physical Anthropology, 161*(4), 698–704. https://doi.org/10.1002/ajpa.23074
Ma, K., Zhao, X., Li, H., Cao, Y., Li, W., Ouyang, J., Xie, L., & Liu, W. (2018). *Forensic Science International: Genetics, 32*, 88–93. https://doi.org/10.1016/j.fsigen.2017.11.001
Melchionda, F., Stanciu, F., Buscemi, L., Pesaresi, M., Tagliabracci, A., & Turchi, C. (2020). Searching the undetected mtDNA variants in forensic MPS data. *Forensic Science International Genetics*, *49*, 102399. https://doi.org/10.1016/j.fsigen.2020.102399. Epub 2020 Sep 28. PMID: 33038616.
Meng, Z., Lin, X., He, X., Gao, Y., Liu, H., Liu, Y., Zhou, Y., Li, J., Chen, Z., Zhang, S., & Li, Y. (2011). *Construction of the platform for phylogenetic analysis*. Data Driven e-Science.

Miao, X., Shen, Y., Gong, X., Yu, H., Li, B., Chang, L., Wang, Y., Fan, J., Liang, Z., Tan, B., Li, S., & Zhang, B. (2021). *International Journal of Legal Medicine, 135*(3), 709–718. https://doi.org/10.1007/s00414-020-02403-z

Mikkelsen, M., Frank-Hansen, R., Hansen, A. J., & Morling, N. (2014). *Forensic Science International: Genetics, 12*, 30–37. https://doi.org/10.1016/j.fsigen.2014.03.014

Novroski, N. M. M., King, J. L., Churchill, J. D., Seah, L. H., & Budowle, B. (2016). *Forensic Science International: Genetics, 25*, 214–226. https://doi.org/10.1016/j.fsigen.2016.09.007

Pakstis, A. J., Gandotra, N., Speed, W. C., Murtha, M., Scharfe, C., & Kidd, K. K. (2021). *Human Genetics, 140*(12), 1753–1773. https://doi.org/10.1007/s00439-021-02382-0

Pang, J. B., Rao, M., Chen, Q. F., Ji, A. Q., Zhang, C., Kang, K. L., Wu, H., Ye, J., Nie, S. J., & Wang, L. (2020). *Scientific Reports, 10*(1). https://doi.org/10.1038/s41598-020-58980-x

Park, S., Cho, S., Seo, H. J., Lee, J. H., Kim, M. Y., & Lee, S. D. (2017). *Journal of Korean Medical Science, 32*(4), 587–592. https://doi.org/10.3346/jkms.2017.32.4.587

Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T., & Reich, D. (2012). Ancient admixture in human history. *Genetics, 192*(3),1065–1093. https://doi.org/10.1534/genetics.112.145037. Epub 2012 Sep 7. PMID: 22960212; PMCID: PMC3522152.

Peck, M. A., Brandhagen, M. D., Marshall, C., Diegoli, T. M., Irwin, J. A., & Sturk-Andreaggi, K. (2016). *Forensic Science International: Genetics, 24*, 103–111. https://doi.org/10.1016/j.fsigen.2016.06.003

Peng, D., Zhang, Y., Ren, H., Li, H., Li, R., Shen, X., Wang, N., Huang, E., Wu, R., & Sun, H. (2020). *Scientific Reports, 10*(1). https://doi.org/10.1038/s41598-020-69137-1

Pereira, V., Mogensen, H. S., Børsting, C., & Morling, N. (2017). *Forensic Science International: Genetics, 28*, 138–145. https://doi.org/10.1016/j.fsigen.2017.02.013

Phillips, C., McNevin, D., Kidd, K. K., Lagacé, R., Wootton, S., de la Puente, M., Freire-Aradas, A., Mosquera-Miguel, A., Eduardoff, M., Gross, T., Dagostino, L., Power, D., Olson, S., Hashiyada, M., Oz, C., Parson, W., Schneider, P. M., Lareu, M. V., & Daniel, R. (2019). *Forensic Science International: Genetics, 42*, 213–226. https://doi.org/10.1016/j.fsigen.2019.06.022

Pritchard, J. K., Stephens, M. J., & Donnelly, P. J. (2000). Inference of population structure using multilocus genotype data. *Genetics, 155*(2), 945–959. https://doi.org/10.1093/genetics/155.2.945

Reich, D., Price, A. L., & Patterson, N. (2008). Principal component analysis of genetic data. *Nature Genetics, 40*(5), 491–492. https://doi.org/10.1038/ng0508-491

Rosenberg, N. A. (2004). Distruct: A program for the graphical display of population structure. *Molecular Ecology Notes, 4*(1), 137–138. https://doi.org/10.1046/j.1471-8286.2003.00566.x

Salvador, J. M., Apaga, D. L. T., Delfin, F. C., Calacal, G. C., Dennis, S. E., & De Ungria, M. C. A. (2018). *Forensic Science International: Genetics, 36*, e8–e12. https://doi.org/10.1016/j.fsigen.2018.06.008

Salvoro, C., Faccinetto, C., Zucchelli, L., Porto, M., Marino, A., Occhi, G., de los Campos, G., & Vazza, G. (2019). *Forensic Science International: Genetics, 40*, 192–200. https://doi.org/10.1016/j.fsigen.2019.03.008

Seo, S. B., King, J. L., Warshauer, D. H., Davis, C. P., Ge, J., & Budowle, B. (2013). Single nucleotide polymorphism typing with massively parallel sequencing for human identification. *International Journal of Legal Medicine, 127*(6), 1079–1086. https://doi.org/10.1007/s00414-013-0879-7

Servedio, M. R., Brandvain, Y., Dhole, S., Fitzpatrick, C. L., Goldberg, E. E., Stern, C. A., Van Cleve, J., & Yeh, D. J. (2014). Not just a theory–the utility of mathematical models in evolutionary biology. *PLoS Biology, 12*(12). e1002017. https://doi.org/10.1371/journal.pbio.1002017.

Silva, D. S. B. S., Scheible, M. K., Bailey, S. F., Williams, C. L., Allwood, J. S., Just, R. S., Schuetter, J., Skomrock, N., Minard-Smith, A., Barker-Scoggins, N., Eichman, C., Meiklejohn, K., & Faith, S. A. (2020). *Forensic Science International: Genetics, 48*, 102311. https://doi.org/10.1016/j.fsigen.2020.102311

Stoljarova, M., King, J. L., Takahashi, M., Aaspõllu, A., & Budowle, B. (2016). Whole mitochondrial genome genetic diversity in an Estonian population sample. *International Journal of Legal Medicine, 130*(1), 67–71. https://doi.org/10.1007/s00414-015-1249-4. Epub 2015 Aug 20. PMID: 26289416.

Strobl, C., Churchill Cihlar, J., Lagacé, R., Wootton, S., Roth, C., Huber, N., Schnaller, L., Zimmermann, B., Huber, G., Lay Hong, S., Moura-Neto, R., Silva, R., Alshamali, F., Souto, L., Anslinger, K., Egyed, B., Jankova-Ajanovska, R., Casas-Vargas, A., Usaquén, W., … Parson, W. (2019). Evaluation of mitogenome sequence concordance, heteroplasmy detection, and haplogrouping in a worldwide lineage study using the Precision ID mtDNA Whole Genome Panel. *Forensic Science International Genetics, 42*, 244–251. https://doi.org/10.1016/j.fsigen.2019.07.013. Epub 2019 Jul 23. PMID: 31382159.

Sun, S., Liu, Y., Li, J., Yang, Z., Wen, D., Liang, W., Yan, Y., Yu, H., Cai, J., & Zha, L. (2020). *Forensic Science International: Genetics, 46*, 102255. https://doi.org/10.1016/j.fsigen.2020.102255

Ta, M. T. A., Nguyen, N. N., Tran, D. M., Nguyen, T. H., Vu, T. A, Le, D. T., Le, P. T., Do, T. T. H., Hoang, H., & Chu, H. H. (2021). Massively parallel sequencing of human skeletal remains in Vietnam using the precision ID mtDNA control region panel on the Ion S5™ system. *International Journal of Legal Medicine*, *135*(6), 2285–2294. https://doi.org/10.1007/s00414-021-02649-1. Epub 2021 Jul 1. PMID: 34196785.

Tao, R., Wang, S., Chen, A., Xia, R., Zhang, X., Yang, Q., Qu, Y., Zhang, S., & Li, C. (2021). Parallel sequencing of 87 STR and 294 SNP markers using the prototype of the SifaMPS panel on the MiSeq FGx system. *Forensic Science International*, *52*, 102490. https://doi.org/10.1016/j.fsigen.2021.102490. Epub 2021 Mar 4. PMID: 33689955.

Taylor, C. R., Kiesler, K. M., Sturk-Andreaggi, K., Ring, J. D., Parson, W., Schanfield, M., Vallone, P. M., & Marshall, C. (2020). Platinum-quality mitogenome haplotypes from United States populations. *Genes (Basel), 11*(11), 1290. https://doi.org/10.3390/genes11111290. PMID: 33138247; PMCID: PMC7716222.

van der Gaag, K. J., de Leeuw, R. H., Laros, J. F. J., den Dunnen, J. T., & de Knijff, P. (2018). *Forensic Science International: Genetics, 35*, 169–175. https://doi.org/10.1016/j.fsigen.2018.05.008

Wang, M., Wang, Z., He, G., Liu, J., Wang, S., Qian, X., Lang, M., Li, J., Xie, M., Li, C., & Hou, Y. (2019). *Forensic Science International: Genetics, 38*, 70–76. https://doi.org/10.1016/j.fsigen.2018.10.009

Wang, M., Wang, Z., He, G., Wang, S., Zou, X., Liu, J., Wang, F., Ye, Z., & Hou, Y. (2020). *Forensic Science International: Genetics, 44*. https://doi.org/10.1016/j.fsigen.2019.102197

Wang, Z., Wang, L., Liu, J., Ye, J., & Hou, Y. (2020). *Electrophoresis, 41*(3–4), 194–201. https://doi.org/10.1002/elps.201900278

Wang, Z., Qin, L., Liu, J., Jiang, L., Zou, X., Chen, X., Song, F., Dai, H., & Hou, Y. (2022). *Forensic Science International: Genetics, 57*. https://doi.org/10.1016/j.fsigen.2021.102657

Warshauer, D. H., Churchill, J. D., Novroski, N., King, J. L., & Budowle, B. (2015). *Genomics, Proteomics & Bioinformatics, 13*(4), 250–257. https://doi.org/10.1016/j.gpb.2015.08.001

Wen, D., Sun, S., Liu, Y., Li, J., Yang, Z., Kureshi, A., Fu, Y., Li, H., Jiang, B., Jin, C., Cai, J., & Zha, L. (2021). *Electrophoresis, 42*(9–10), 1115–1126. https://doi.org/10.1002/elps.202000341

Wendt, F. R., King, J. L., Novroski, N. M. M., Churchill, J. D., Ng, J., Oldt, R. F., McCulloh, K. L., Weise, J. A., Smith, D. G., Kanthaswamy, S., & Budowle, B. (2017). *Forensic Science International: Genetics, 28*, 146–154. https://doi.org/10.1016/j.fsigen.2017.02.014

Wisner, M., Erlich, H., Shih, S., & Calloway, C. (2021). Resolution of mitochondrial DNA mixtures using a probe capture next generation sequencing system and phylogenetic-based software. *Forensic Science International Genetics*, *53*, 102531. https://doi.org/10.1016/j.fsigen.2021.102531. Epub 2021 May 19. PMID: 34052577.

Wu, L., Chu, X., Zheng, J., Xiao, C., Zhang, Z., Huang, G., Li, D., Zhan, J., Huang, D., Hu, P., & Xiong, B. (2019). *Forensic Science International: Genetics, 42*, 227–234. https://doi.org/10.1016/j.fsigen.2019.07.006

Wu, H., Ji, A. C., Liu, Y. C., Kang, K. L., Zhang, C., Li, Z. W., Ji, A. Q., Ye, J., Nie, S. J., & Wang, L. (2020). *Electrophoresis, 41*(23), 2029–2035. https://doi.org/10.1002/elps.202000093

Wu, R., Li, H., Li, R., Peng, D., Wang, N., Shen, X., & Sun, H. (2021). *International Journal of Legal Medicine, 135*(4), 1137–1149. https://doi.org/10.1007/s00414-020-02483-x

Wu, Riga, Chen, H., Li, R., Zang, Y., Shen, X., Hao, B., Wang, Q., & Sun, H. (2021). *Forensic Science International, 325*, 110875. https://doi.org/10.1016/j.forsciint.2021.110875

Xie, T., Shen, C., Liu, C., Fang, Y., Guo, Y., Lan, Q., Wang, L., Ge, J., Zhou, Y., Wen, S., Yang, Q., & Zhu, B. (2020). *Journal of Human Genetics, 65*(5), 461–468. https://doi.org/10.1038/s10038-020-0725-y

Yang, J., Chen, J., Ji, Q., Yu, Y., Li, K., Kong, X., Xie, S., Zhan, W., Mao, Z., Yu, Y., Li, D., Chen, P., & Chen, F. (2022). *Forensic Science International: Genetics, 56*, 102600. https://doi.org/10.1016/j.fsigen.2021.102600

Zhang, S., Bian, Y., Chen, A., Zheng, H., Gao, Y., Hou, Y., & Li, C. (2017a). *Forensic Science International: Genetics, 27*, 50–57. https://doi.org/10.1016/j.fsigen.2016.12.003

Zhang, S., Bian, Y., Chen, A., Zheng, H., Gao, Y., Hou, Y., & Li, C. (2017b). *Forensic Sciences Research, 2*(1), 26–33. https://doi.org/10.1080/20961790.2017.1281011

Zhang, S., Niu, Y., Bian, Y., Dong, R., Liu, X., Bao, Y., Jin, C., Zheng, H., & Li, C. (2018). *Scientific Reports, 8*(1). https://doi.org/10.1038/s41598-018-24495-9

Zhang, Q. X., Yang, M., Pan, Y. J., Zhao, J., Qu, B. W., Cheng, F., Yang, Y. R., Jiao, Z. P., Liu, L., & Yan, J. W. (2018). Development of a massively parallel sequencing assay for investigating sequence polymorphisms of 15 short tandem repeats in a Chinese Northern Han population. *Electrophoresis*, *39*(21), 2725–2731. https://doi.org/10.1002/elps.201800071. Epub 2018 Jun 4. PMID: 29772597.

Zhao, X., Ma, K., Li, H., Cao, Y., Liu, W., Zhou, H., & Ping, Y. (2015). *Forensic Science International: Genetics, 19*, 192–196. https://doi.org/10.1016/j.fsigen.2015.06.012

Zhao, X., Li, H., Wang, Z., Ma, K., Cao, Y., & Liu, W. (2016). *Forensic Science International: Genetics, 25*, 34–38. https://doi.org/10.1016/j.fsigen.2016.07.014

Zhao, G.-B., Ma, G.-J., Zhang, C., Kang, K.-L., Li, S.-J., & Wang, L. (2021). *Forensic Science International: Genetics, 55*, 102580. https://doi.org/10.1016/j.fsigen.2021.102580

Zhao, X., Fan, Y., Zeye, M. M. J., He, W., Wen, D., Wang, C., Li, J., & Hua, Z. (2022). *International Journal of Legal Medicine, 136*(1), 43–53. https://doi.org/10.1007/s00414-021-02719-4

Zhou, Y., Guo, F., Yu, J., Liu, F., Zhao, J., Shen, H., Zhao, B., Jia, F., Sun, Z., Song, H., & Jiang, X. (2016). *Forensic Science International: Genetics, 22*, 11–21. https://doi.org/10.1016/j.fsigen.2016.01.004

Zhu, J., Lv, M., Zhou, N., Chen, D., Jiang, Y., Wang, L., He, W., Peng, D., Li, Z., Qu, S., Wang, Y., Wang, H., Luo, H., An, G., Liang, W., & Zhang, L. (2019). *Forensic Science International: Genetics, 39*, 1–7. https://doi.org/10.1016/j.fsigen.2018.11.005

Zou, X., He, G., Liu, J., Jiang, L., Wang, M., Chen, P., Hou, Y., & Wang, Z. (2022). *Forensic Science International: Genetics, 58*. https://doi.org/10.1016/j.fsigen.2022.102687

This page intentionally left blank

SECTION 2

Useful applications of NGS-based forensic DNA analysis

This page intentionally left blank

CHAPTER 9

Applications of NGS in analysis of autosomal STRs

Hannah Berry
Towson University, Towson, MD, United States

Introduction

Short tandem repeats (STRs) are regions of DNA that repeat a number of times, and that number can vary. These repeats typically contain between two and six base pairs (Wyner et al., 2020). Autosomal STRs are found in the 22 pairs of autosomal chromosomes that humans have. These chromosomes do not determine a person's sex (What is STR Analysis? National Institute of Justice, 2011). Because the number of repeats STRs have is variable, they are often used in forensics to distinguish individuals (Wyner et al., 2020). Several kits are commercially produced to amplify autosomal STRs as well as STRs found on both the X and Y chromosomes using both next-generation sequencing (NGS) and capillary electrophoresis (CE). NGS provides a more detailed way of analyzing the STRs because it can determine both the length and sequence of the loci instead of only the length determined by CE. One way of analyzing the STRs is by performing population studies that show variability in sequence and length by population. Using NGS methods, the sequences of both flanking regions and the repeat region of STR loci can be determined, and certain sequences can be more prevalent in certain populations (Kwon et al., 2021).

Next-generation sequencing

NGS, otherwise known as massively parallel sequencing (MPS), is a method of determining the DNA sequence(s) found in a sample. NGS is a high-throughput technology that allows for the sequences of multiple DNA templates to be read at the same time, in parallel. Whole genome sequencing can be performed using NGS technologies to amplify and sequence the entire genome. Targeted sequencing can also be completed to determine the sequence of specific regions on the DNA template. Multiple methods of NGS exist, including sequencing-by-synthesis, pyrosequencing, and ion detection (Bruijns et al., 2018).

Sequencing-by-synthesis using reversible terminators, created by Illumina, begins with adapters being ligated to the template DNA, and those adapters will bind to complementary oligonucleotides found on the flow cell. Adapters on both ends of the DNA

Next Generation Sequencing (NGS) Technology in DNA Analysis
ISBN 978-0-323-99144-5, https://doi.org/10.1016/B978-0-323-99144-5.00012-3
© 2024 Elsevier Inc.
All rights reserved.

strand will bind to the complementary sequences, forming a bridge. Bridge amplification then occurs, and clusters of the same DNA sequence form in the same area, with multiple clusters of different DNA sequences forming. Once denatured, dideoxynucleotides with fluorescent labels are added to complement the DNA strand. Each type of ddNTP has its own fluorescent label, so as the bases are added, the unique fluorescence is detected (Bruijns et al., 2018). This allows the sequence of the DNA strand to be determined.

Pyrosequencing, used in the Roche 454 technology, is another sequencing-by-synthesis method. As a complementary base is incorporated into the complementary DNA strand, a conversion from pyrophosphate to adenosine triphosphate (ATP) occurs. This ATP drives the conversion to oxyluciferin after the addition of luciferin. When this occurs, light is emitted. So, when the complementary nucleotide is added to the strand being synthesized, light will be detected (Bruijns et al., 2018). The sequence of the template DNA strand can be determined by adding only one dNTP at a time.

Ion detection, introduced by Ion Torrent, detects changes in pH instead of light and fluorescence like the previous two methods described. When pyrophosphate is cleaved after the addition of a nucleotide to the DNA strand, a proton is released. This causes a voltage change. If one specific nucleotide is added to the DNA being sequenced and the voltage changes, then that nucleotide is incorporated (Bruijns et al., 2018). As different nucleotides are added to the reaction, these can be monitored to figure out the sequence of the DNA. These NGS technologies, along with others, have been applied in a variety of ways, one of which is the analysis of autosomal STRs.

Application of NGS in autosomal STR analysis

NGS is a newer method for the analysis of STRs and SNPs that can be utilized in the field of forensic science. Using STR profiles obtained from NGS methods, individuals can be identified, whether they are victims or suspects. Samples received by technicians and analysts for forensic analysis are often small and contain low-quality and/or low-quantity DNA. NGS has been shown to be suitable for the analysis of STRs on DNA that is degraded or found in low quantities (Bruijns et al., 2018). The ability of this technology to be used to amplify short fragments is an advantage over the traditionally used CE method.

NGS has been used to help identify victims of World War II after the discovery of skeletal remains in the Babna Gora municipality in Slovenia. Using the Precision ID GlobalFiler NGS STR Panel, which includes 31 autosomal STRs, genotyping was only successful for one out of three incomplete skeletons. To attempt to identify the individual, samples were also taken from the presumed niece and nephew of the individual. After analysis, the posterior probability was calculated to be 99.99986%, meaning that the victim sample was most likely related to the presumed relatives' samples rather than related to an unrelated individual (Pajnič et al., 2019). So, NGS can be useful for the

identification of victims by comparing victim samples to family samples and completing the analysis of autosomal STRs.

A challenge for identifications using STR profiles obtained after CE is in cases involving monozygotic or identical twins. After CE, the STR profiles of identical twins are the same, so the two individuals cannot be differentiated. This can be problematic when working with forensic evidence from crime scenes or determining if an individual is the father of a child for paternity testing. To be able to differentiate between identical twins, sequencing DNA and locating mutations are necessary. Weber-Lehmann et al. used Illumina's technology to locate single nucleotide polymorphisms of two identical twins and the child of one identical twin. They were able to locate many sites with single base pair polymorphisms, some including sites on chromosomes 4, 6, and 14, that were different between the two twins and the same in the parent/child profiles (Weber-Lehmannn et al., 2014).

In addition to being able to differentiate between identical twins, NGS has also been shown to differentiate the genetic profiles of nonrelated individuals in samples with mixed DNA contributed by multiple people. Mixed samples are often collected in forensic cases, containing mixtures of suspects' and victims' DNA. Momota et al. (2021) were able to determine whether using NGS would be a beneficial tool when analyzing samples with more than one DNA contributor. They created different ratios of two-person mixtures. After performing NGS, they found that differences in the STR sequences could be determined up to a mixing ratio of 1:5 (Momota et al., 2021). This indicates that autosomal STR profiles can be separated out for two-person mixtures, which can be useful in forensic contexts when evidence has multiple DNA contributors.

As the utility of NGS in forensics becomes more apparent, developmental validations are being completed for the instruments and software programs that analyze STRs as well as SNPs using NGS. Developmental validations are necessary for instruments, software, and reagents needed for forensic analysis as they determine the reliability of the test. Many elements are included within a developmental validation, including studies on reproducibility, accuracy, and sensitivity, among others (National Commission on Forensic Science: Views of the Commission: Validation of Forensic Science Methodology National Institute of Standards and Technology, 2/29/16). One such validation was completed for the MiSeq FGx Forensic Genomics System. The ForenSeq DNA Signature Prep Kit, included in Table 1, is part of this system. Sensitivity studies test dilutions of DNA to determine the percentage of allele calls made with varying amounts of DNA. Reproducibility studies determine whether the method produces the same results for multiple analysts completing the same method. Accuracy studies establish whether the results of the analysis are correct. These tests were all performed with the MiSeq FGx Forensic Genomics System. The group found that for the sensitivity study, a range of between 1 ng and 62.5 pg resulted in no allele loss. The reproducibility test, completed by five analysts, was used to determine accuracy in addition to precision. The results showed

that for primer set A, there was 100% accuracy and 99.89% precision. For primer set B, the results were 100% for both accuracy and precision (Jäger et al., 2017). These findings show that the MiSeq FGx Forensic Genomics System, utilizing NGS, can be reliable for forensic analysis of autosomal STRs.

When analyzing STRs, traditional methods involve the use of CE to develop the STR profiles. As NGS is being introduced into more laboratories, it has been demonstrated that NGS can improve information gained from profiles when compared to CE. These two methods were compared in a study by Revoir et al. (2019). Fifty-three DNA samples were used and amplified using the DNA Signature Prep Kit for NGS and the Promega PowerPlex Fusion 6C kit for CE. These two kits have 26 targeted loci in common. After gathering the results, there was 99.3% concordance between the NGS kit and the CE kit initially. When the results in the Universal Analysis Software for the MiSeq were reviewed, the concordance percentage went up to 99.8%. This was due to calls made subthreshold. For the remaining disagreements between the two kits, they were found to be because of missing allelic data at different loci. They determined that this was possibly due to inefficient amplification or dropout, particularly at the Penta E loci. In addition to determining concordance between the methods, they also looked at how allele counts changed. The change in allele count is due to NGS being able to distinguish allele sequences instead of allele determination by size using CE. Seventeen out of twenty-three loci increased in the number of alleles when determining alleles by sequence. The largest increase in allele number was exhibited by the D12S391 locus, which gained an additional 20 alleles (Revoir et al., 2019). This study demonstrates the ability of NGS to be a beneficial method for the analysis of STRs. The concordance of allele calls between the NGS and CE was greater than 99%; however, the NGS provides a way of getting sequence information that can distinguish different alleles by more than just size.

Comparing loci amplified using different kits for NGS applications, including two kits for CE applications

Table 9.1 shows the list of autosomal STR loci that are targeted for the ForenSeq DNA Signature Prep Kit, Precision ID GlobalFiler NGS STR Panel v2, PowerSeq 46GY System, Investigator 24Plex QS Kit, and the PowerPlex Fusion System. The first three kits are used for NGS applications, while the last two kits are used for analysis using CE. Based on the number of autosomal STR loci targeted for each kit, the kits for NGS applications detect the most loci when compared to those for use with CE. Loci such as D4S2408 and D6S1043 are included in more than one kit for NGS but not in the kits for CE. Only one locus, SE33, is included in the MainstAY SE available-for-purchase NGS-based multiplex kit discussed here. The SE33 locus is included when amplifying DNA using Qiagen's Investigator 24Plex kit.

Table 9.1 Loci Amplified for 4 NGS kits and 2 CE kits.

	Verogen ForenSeq MainstAY SE kit (Forensic Mainstay Kit Reference Guide) (NGS)	Verogen ForenSeq DNA Signature Prep kit (ForenSeq DNA Signature Prep Reference) (NGS)	Precision ID GlobalFiler NGS STR panel v2 (Precision ID GlobalFilerTM NGS STR, 1612) (NGS)	Promega PowerSeq 46GY system (PowerSeq 4GY system Technical Manual) (NGS)	Qiagen Investigator 24Plex QS (Investigator 24plex QS Handbook) (CE)	Promega PowerPlex Fusion system (PowerPlex Fusion system for use) (CE)
D1S1656	√	√	√	√	√	√
TPOX	√	√	√	√	√	√
D2S441	√	√	√	√	√	√
D2S1338	√	√	√	√	√	√
D3S1358	√	√	√	√	√	√
D4S2408	√	√	√			
FGA	√	√	√	√	√	√
D5S818	√	√	√	√	√	√
CSF1PO	√	√	√	√	√	√
D6S1043	√	√	√			
D7S820	√	√	√	√	√	√
D8S1179	√	√	√	√	√	√
D9S1122	√	√				
D10S1248	√	√	√	√	√	√
TH01	√	√	√	√	√	√
vWA	√	√	√	√	√	√
D12S391	√	√	√	√	√	√
D13S317	ü	√	√	√	√	√
PentaE	√	√	√	√		√
D16S539	√	√	√	√	√	√
D17S1301	√	√				
D18S51	√	√	√	√	√	√
D19S433	√	√	√	√	√	√

Continued

Table 9.1 Loci Amplified for 4 NGS kits and 2 CE kits.—cont'd

	Verogen ForenSeq MainstAY SE kit (Forensic Mainstay Kit Reference Guide) (NGS)	Verogen ForenSeq DNA Signature Prep kit (ForenSeq DNA Signature Prep Reference) (NGS)	Precision ID GlobalFiler NGS STR panel v2 (Precision ID GlobalFilerTM NGS STR, 1612) (NGS)	Promega PowerSeq 46GY system (PowerSeq 4GY system Technical Manual) (NGS)	Qiagen Investigator 24Plex QS (Investigator 24plex QS Handbook) (CE)	Promega PowerPlex Fusion system (PowerPlex Fusion system for use) (CE)
D20S482	√	√				
D21S11	√	√	√	√	√	√
PentaD	√	√	√	√		√
D22S1045	√	√	√	√	√	√
D1S1677			√			
D2S1776			√			
D3S4529			√			
D5S2800			√			
D6S474			√			
D12ATA63			√			
D14S1434			√			
SE33	√				√	

When comparing the NGS STR kits with each other, the Promega PowerPlex 46GY System amplifies the least number of loci, totaling 22. This kit includes all twenty of the core Combined DNA Index System (CODIS) loci, as well as those included in the European Standard Set (ESS) (Switching to 20 Core CODIS, 1210; Core STR Loci Used in). Verogen's ForenSeq DNA Signature Prep Kit amplifies a total of 27 loci. The loci this kit targets also include the expanded set of CODIS loci and the ESS loci. Verogen's MainstAY SE kit, which is in the beta testing stage as of June 2022, includes all of the loci amplified in the ForenSeq DNA Signature Prep Kit as well as SE33. The Precision ID GlobalFiler NGS STR Panel v2 amplifies the most loci when compared to the other two kits, amplifying 31 different autosomal STRs. CODIS and ESS loci are targeted with this kit as well.

Population studies of autosomal STRs using NGS

Different population groups in the world can have differences in allele frequencies for particular STRs as well as variance in the DNA sequence of the STRs that have been amplified. Kwon et al. (2021) used an in-house panel for NGS and analyzed samples obtained from four groups: African American, Caucasian, Hispanic, and Korean. A total of 25 autosomal STRs were analyzed by this group. One STR that was highlighted was SE33, a repeat that is highly polymorphic in terms of both length and sequence. It was analyzed due to a lack of data on its sequence polymorphisms when compared to other STR markers. After creating the MPS panel, MPS was completed using a PCR-based method for creating the library. When analyzing SE33, which is characterized by [CTTT] repeats, however, differences in the repeats were noted. Isoalleles were also identified and counted for both repeat and flanking regions. Allele frequencies were calculated based on the sequences in the 5′ and 3′ flanking regions of SE33. For the African American and Hispanic samples that were tested, the highest percentage of alleles observed had [TTTC] preceding the repeat region, while for Caucasians and Koreans, [CTTC] was the most observed sequence at that location. The most SE33 sequence variations were exhibited in the African American population, with nine different sequence variations, while the least number of variations was found in the Hispanic population. The Hispanic population only exhibited three variations. Seven total positions of polymorphisms were found in SE33 (Kwon et al., 2021).

The SE33 locus was also able to be analyzed using the Verogen ForenSeq MainstAY SE kit, which was being beta-tested at Towson University. The researchers there tested samples from Caucasian and African American populations. They found, of the three Caucasian samples tested, all had isoalleles at the SE33 locus. In one sample, however, the isoallele was detected below threshold. When compared to the African American population, the number of samples with isoalleles observed dropped down to 4 out of 19 tested samples (Noelle Neff, unpublished results).

Population studies and the data they generate allow for statistical calculations to be performed in forensic cases. These statistics determine the likelihood that an individual is a contributor to DNA found on a piece of evidence. One group, Gettings et al. (2018), sought to increase the amount of available allele frequency data for different population groups in order to determine likelihood statistics. They targeted 27 different autosomal loci, and the frequencies are reported with high confidence. African American, Caucasian, Asian, and Hispanic population groups were represented in the samples. Some of the STR loci targeted include D12S391, D18S51, and D4S2408. The analyses were performed using two different software types: Universal Analysis Software by Illumina and a pipeline derived from STRait Razor 2.0. After the analyses were completed, the group found that 16/27 loci had sequence polymorphisms in their flanking regions, totaling 40 polymorphisms. The majority of these variations were single-nucleotide polymorphisms, but insertions and deletions were also found. Sequence motifs were determined, and the frequency within a particular population was calculated for the motifs. These frequencies varied by population and motif. For example, the data showed that at the D12S391 loci, the [AGAT]n [AGAC]n AGAT motif most frequently occurred in Asian populations, with a 90.72% allele frequency in that population. In the Caucasian population, this number was lower, with a 66.07% frequency. At the D13S317 loci, the [TATC]n motif was most common in the African American population at 69.74%. The same motif had a frequency of 41.75% in the Asian population (Gettings et al., 2018). Reporting of these population frequencies helps with statistical calculations used in forensics to determine the likelihood of a person contributing DNA to a sample.

Dai et al. (2019) conducted a population study for the Chinese Han people. Analysis of the STRs obtained from DNA in blood samples was completed after NGS. When compared with the number of alleles for a particular locus using CE methods, the number of alleles after NGS increased. This change was due to single nucleotide polymorphisms being recognized in the repeat or flanking region of the STR and differences in the structure of the repeat. Differences in the structure of the repeat were noted for the seven compound loci that were amplified. At least two times the number of alleles at several loci identified using CE were identified using NGS. These loci include D2S1338, D13S317, D2S11, and vWA, among others. Using next-generation sequencing, this group was able to determine the number of alleles for each amplified locus in the Chinese Han community (Dai et al., 2019). A similar result was found for other population groups in another research study.

Illumina's sequencing technology was used by Gettings et al. (2016) when researching variations in the DNA sequences of autosomal STRs. Twenty-two loci were analyzed in this study. DNA samples were used to represent people in the African American, Caucasian, and Hispanic populations. The group found that for 15 of the loci, there was an increase in the number of alleles based on sequence when compared to alleles based on length using NGS. This shows that NGS has a greater ability to discriminate between different alleles than by using CE methods. This is due to its capability to generate information at the sequence level of the DNA. In addition, they were also able to calculate the

probability of identity (PI) for each population group. In general, the PI decreased when calculated based on sequence instead of length. This value varied for each population, though. For example, at the D12S391 loci, the PI was 0.020 for African Americans, 0.019 for Europeans, and 0.028 for Hispanics when calculating PI by sequence. This PI value is the probability that, at a particular locus, two people within the same population have the same genotype (Gettings et al., 2016).

The D21S11 locus was analyzed by Rockenbauer et al. (2014) using samples obtained from 29 paternity cases in Denmark. Not all of the individuals were related; samples used from five of the cases were from unrelated individuals. Sequencing was performed using Roche Diagnostics' GS Junior System, which utilizes the pyrosequencing method. After running CE on the samples, it was indicated that at D21S11, a mutation might have occurred between a parent and the child in 54 confirmed father/mother/child-related samples. NGS was able to suggest which of the alleles coming from the parent had been mutated. The mutations observed in the locus alleles were either single or multiple repeat units being inserted or deleted. After sequencing the unrelated individuals' DNA, 20 alleles for the D21S11 locus were recognized, while only 13 were identified after CE. Differences in sequences for the same allele call were contributed to by variations in the number of subrepeats (Rockenbauer et al., 2014). So, while the length of the alleles may be the same in a case sample, the actual sequence may be different after looking at NGS data.

Conclusion

NGS is a newer technology that is able to analyze STRs, including autosomal STRs. NGS can determine the DNA sequences as well as the length of the STR loci. Since the sequence can be discovered, this has an advantage over the STR analysis capabilities of CE, which can distinguish STRs by size. Many population studies have been performed to calculate population allele frequencies. These values are used in forensics to determine the likelihood that a person's DNA matches that found on evidence as that it came from a different person. Analysis of autosomal STRs using NGS has also been shown to be useful for separating DNA mixtures as well as detecting differences between monozygotic twins. NGS has many applications for the analysis of autosomal STRs; however, its usefulness expands beyond autosomal STRs.

References

Bruijns, B., Tiggelaar, R., & Gardeniers, H. (2018). Massively parallel sequencing techniques for forensics: A review. *Electrophoresis, 39*(21), 2642−2654. https://doi.org/10.1002/elps.201800082

Core str loci used in human identity testing. NIST. https://strbase.nist.gov/coreSTRs.htm.

Dai, W., Pan, Y., Sun, X., Wu, R., Li, L., & Yang, D. (2019). High polymorphism detected by massively parallel sequencing of autosomal STRs using old blood samples from a Chinese Han population. *Scientific Reports, 9*(1), 18959. https://doi.org/10.1038/s41598-019-55282-9

ForenSeq DNA signature prep reference guide. Verogen. https://www.qiagen.com/us/resources/download.aspx?id=f9829e77-412e-47aa-bd53-12fb4add19e8&lang=en.

Forensic mainstAY kit reference guide. Verogen.

Gettings, K. B., Borsuk, L. A., Steffene, C. R., Kiesler, K. M., & Vallone, P. M. (2018). Sequence-based U.S. population data for 27 autosomal STR loci. *Forensic Science International: Genetics, 37*, 106–115. https://doi.org/10.1016/j.fsigen.2018.07.013

Investigator 24plex QS handbook. Qiagen. https://www.qiagen.com/us/resources/download.aspx?id=debe09ab-5483-478b-aeb3-e5c128e78a92&lang=en.

Gettings, K. B., Kiesler, K. M., Faith, S. A., Montano, E., Baker, C. H., Young, B. A., Guerrieri, R. A., & Vallone, P. M. (2016). Sequence variation of 22 autosomal STR loci detected by next generation sequencing. *Forensic Science International: Genetics, 21*, 15–21. https://doi.org/10.1016/j.fsigen.2015.11.005

Jäger, A. C., Alvarez, M. L., Davis, C. P., Guzmán, E., Han, Y., Way, L., Walichiewicz, P., Silva, D., Pham, N., Caves, G., Bruand, J., Schlesinger, F., Pond, S., Varlaro, J., Stephens, K., & Holt, C. (2017). Developmental validation of the MiSeq FGx forensic genomics system for targeted next generation sequencing in forensic DNA casework and database laboratories. *Forensic Science International: Genetics, 28*, 52–70. https://doi.org/10.1016/j.fsigen.2017.01.011

Kwon, Y. L., Kim, B. M., Lee, E. Y., & Shin, K. J. (2021). Massively parallel sequencing of 25 autosomal STRs including SE33 in four population groups for forensic applications. *Scientific Reports, 11*(1), 4701. https://doi.org/10.1038/s41598-021-82814-z

Momota, F., Tsuji, A., Ishiko, A., & Ikeda, N. (2021). Examination of the usefulness of next-generation sequencing in mixed DNA samples. *Journal of Legal Medicine, 51*, 101874. https://doi.org/10.1016/j.legalmed.2021.101874

National Commission on Forensic Science: Views of the Commission: Validation of Forensic Science Methodology, National Institute of Standards and Technology, 2/29/16, https://www.justice.gov/archives/ncfs/page/file/831546/download.

Noelle Neff, Towson university, Unpublished MainstAY SE Results.

Pajnič, I. Z., Pogorelc, B. G., & Zupanc, T. (2019). Next generation sequencing technology in Second World War victim identification. *Forensic Science International: Genetics Supplement Series, 7*(1), 123–125. https://doi.org/10.1016/j.fsigss.2019.09.048

Powerplex fusion system for use on the applied biosystems genetic analyzers technical manual. Promega. https://www.promega.com/-/media/files/resources/protocols/technical-manuals/101/powerplex-fusion-system-protocol.pdf?rev=6bfba9f5e6c741a5809a5985cc9840fd&sc_lang=en.

PowerSeq 4GY system technical manual. Promega. https://www.promega.com/-/media/files/resources/protocols/technical-manuals/500/powerseq-46gy-system-protocol-tm522.pdf?rev=b506e6276bee4c52b78f3f9a1b5e70be&sc_lang=en.

Precision id globalfiler ngs str panel v2 with the hid ion s5/hid ion genestudio s5 system application guide. ThermoFisher Scientific. https://www.thermofisher.com/document-connect/document-connect.html?url=https://assets.thermofisher.com/TFS-Assets%2FLSG%2Fmanuals%2FMAN0016129_PrecisionIDSTRIonS5_UG.pdf.

Revoir, A. P., Lancaster, H., & Ames, C. (2019). Initial assessment of NGS as a tool to augment routine CE analysis of STRs. *Forensic Science International: Genetics Supplement Series, 7*(1), 747–749. https://doi.org/10.1016/j.fsigss.2019.10.162

Rockenbauer, E., Hansen, S., Mikkelsen, M., Børsting, C., & Morling, N. (2014). Characterization of mutations and sequence variants in the D21S11 locus by next generation sequencing. *Forensic Science International: Genetics, 8*(1), 68–72. https://doi.org/10.1016/j.fsigen.2013.06.011

Switching to 20 core CODIS loci and the impact on SAKI testing. RTI International. https://www.sakitta.org/resources/docs/12104SAKIImpctSAKTstCoreLociv3.pdf.

Weber-Lehmannn, J., Schilling, E., Gradl, G., Richter, D. C., Wiehler, J., & Rolf, B. (2014). Finding the needle in the haystack: Differentiating "identical" twins in paternity testing and forensics by ultra-deep next generation sequencing. *Forensic Science International: Genetics, 9*, 42–46. https://doi.org/10.1016/j.fsigen.2013.10.015

Wyner, N., Barash, M., & McNevin, D. (2020). Forensic autosomal short tandem repeats and their potential association with phenotype. *Frontiers in Genetics, 11*, Article 884. https://doi.org/10.3389/fgene.2020.00884

What is STR analysis?. (2011). Washington, D.C.: National Institute of Justice. https://nij.ojp.gov/topics/articles/what-str-analysis.

CHAPTER 10

Differentiating monozygotic twins using NGS

Cassandra Skrant, Ilya Grabylnikov, Kelly M. Elkins and Cynthia B. Zeller
Department of Chemistry, Forensic Science Program, Towson University, Towson, MD, United States

Introduction to DNA analysis methods

The conventional method of individualizing suspects in most forensic investigations involves short tandem repeat (STR) typing with capillary electrophoresis (CE). Although STR typing is an effective method for identifying individuals, it is not capable of differentiating monozygotic twins. Monozygotic twins come from the same fertilized eggs. As a result, they have identical genomes and identical STR allele length profiles. Therefore, current research has started to focus on methods that can be used to discriminate between monozygotic twins. Multiple different methods to discriminate between monozygotic twins have been attempted, including sequencing the mitochondrial genome, sequencing the nuclear genome, microRNA (miRNA) profiling, and comparing DNA methylation, or the methylome.

The new application of next-generation sequencing (NGS) has allowed for this expansion in the possibilities of differentiating monozygotic twins. The high throughput of NGS allows for a large portion of DNA to be sequenced in a shorter amount of time. For mitochondrial genome sequencing, the entire chromosome can be sequenced as opposed to just the hypervariable regions. Sequencing the entire mitochondrial genome allows more differences within the genome to be located, which is useful for differentiating monozygotic twins. For analyzing the mitochondrial and nuclear genomes, the flanking regions can be important to identify sequence variations. Sequence variations have been found in flanking regions, which can be an important area of study for monozygotic twins because sequence variations can be few and difficult to locate (Fig. 10.1).

Mitochondrial genome sequencing

Recent research has focused on detecting differences in monozygotic twins by using NGS to sequence the mitochondrial genome. One reason for the interest in the mitochondrial genome is that it has fewer DNA repair mechanisms than the nuclear genome, which leads to a higher mutation rate and substitution rate (Wang et al., 2015). These characteristics make it more likely that there will be differences between monozygotic twins. In addition, it has a higher copy number and a smaller genome size, which

Next Generation Sequencing (NGS) Technology in DNA Analysis
ISBN 978-0-323-99144-5, https://doi.org/10.1016/B978-0-323-99144-5.00010-X
© 2024 Elsevier Inc.
All rights reserved.

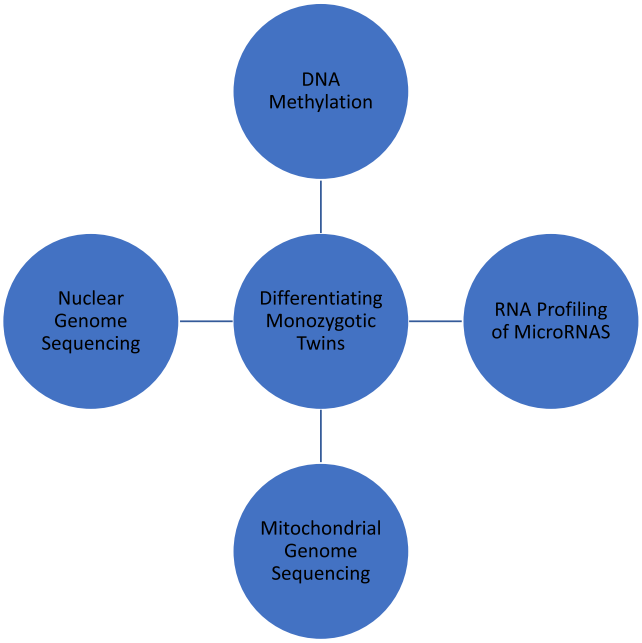


Figure 10.1 Methods for differentiating monozygotic twins.

make it easier to sequence (Yuan et al., 2020). The mitochondrial genomes of six pairs of monozygotic twins were sequenced from blood samples and compared to determine if any heteroplasmy or variation between the monozygotic twins could be determined (Wang et al., 2015). A point heteroplasmy is when there is more than one nucleotide at one position (Wang et al., 2015). The six pairs of monozygotic twins provided blood samples, which were sequenced on the Illumina HiSeq 2000 sequencing system. After sequencing, 11-point heteroplasmies were found (Wang et al., 2015). Three of the heteroplasmic events were in the HV1 or HV2 region, and eight were in coding regions (Wang et al., 2015). The 11 heteroplasmies had varying levels of heterogeneity in all of the monozygotic twin pairs except for one. In addition, there was a single nucleotide variant (nt15301) detected in four out of the six pairs of monozygotic twins (Wang et al., 2015). After comparing the mitochondrial genome in the twin pairs, all of the twin pairs were differentiated by either one of the 11 heteroplasmies found or the nucleotide variant in the mitochondrial genome (Wang et al., 2015).

Sequencing the mitochondrial genome has also been applied to a study of three rape cases and a murderous adultery case that were linked because the STR profiles were identical and both monozygotic twin's STR profiles were consistent with the DNA found at the scenes of the crimes (Yuan et al., 2020). Since two individuals had STR profiles consistent with the evidence, a twin could not be charged until a difference was found. To discriminate between the twins, whole-genome sequencing (WGS) was performed (Yuan et al., 2020). Blood samples from both of the twins were obtained, and the

mitochondrial genomes were sequenced. After sequencing, five single nucleotide polymorphisms (SNPs) were found (Yuan et al., 2020). Amplification refractory mutation system polymerase chain reaction (ARMS-PCR) was used to confirm the SNPs with four primers. After the ARMS-PCR, one somatic mutation in the mitochondrial genome, m.6903 T > C was confirmed (Yuan et al., 2020). There were two different bases at this locus: 2.6% was cytosine and 97.4% was thymine (Yuan et al., 2020). This somatic mutation was found in the nail evidence sample of one twin but not the other. This somatic mutation was used to distinguish the monozygotic twins. The DNA evidence found in both the first and second cases contained the same somatic mutation, so one twin could be identified. In addition, the other twin had two heteroplasmic loci at m.6935C > T and m.6938C > T in the semen samples collected, which were not found in the crime scene evidence samples. Therefore, that twin could be excluded from contributing to the DNA evidence (Yuan et al., 2020). These three differences in the mitochondrial genome were used to differentiate the monozygotic twins.

Nuclear genome sequencing

Along with mutations and heteroplasmies in the mitochondrial genome, SNPs have also been found in the nuclear genome. These are particularly useful in a paternity test involving potential fathers who are monozygotic twins because the SNPs in the nuclear genome are inherited by offspring. Although these polymorphisms are useful, they are extremely rare because the mutation must occur after the splitting of the zygote but as early in development as possible so that it is present in multiple cell types (Weber-Lehmann et al., 2014). If the mutation occurs after the splitting of the zygote and before the separation of the germ layers, it will be present in multiple cell types. If the mutation occurs later in development, it may only be present in sperm cells or other specific cell types. In a study, a paternity test was performed on twin brothers and a child by probing SNPs that arose due to mutations (Weber-Lehmann et al., 2014). Sperm, blood, and mucosa samples were collected from the twin brothers, and blood samples were collected from the child and the mother. Shotgun libraries were prepared and sequenced with an Illumina HiSeq 2000. After sequencing, 12 SNPs were found in only one twin and the child, but seven were eliminated (Weber-Lehmann et al., 2014). The remaining five SNPs were confirmed by Sanger sequencing. The five SNPs were located on chromosomes 4, 6, 11, 14, and 15. Four of these SNPs in the blood samples were also found in the buccal mucosa samples (Weber-Lehmann et al., 2014). These SNPs were used to determine the paternity of the child.

Another feature of the nuclear genome that has been researched for discriminating monozygotic twins is rapidly mutating Y-STRs. Only one research study was found to investigate if rapidly mutating Y-STRs can be used to differentiate monozygotic twins. Samples were collected from 14 pairs of Caucasian-Lebanese male monozygotic twins. For genotyping, the Identifiler Plus and 27-plexY Filer Plus kits were used. The Y-STRs included in the kit were DYS19, DYS385a/b, DYS389I/II, DYS390, DYS391,

DYS392, DYS393, DYS437, DYS438, DYS439, DYS448, DYS456, DYS458, DYS635, GATA H4, DYS460, DYS481, DYS533, DYS387a/b, DYS449, DYS518, DYS570, DYS576, and DYS627 (Romanos & Borjac, 2018). The loci DYS387a/b, DYS449, DYS518, DYS570, DYS576, and DYS627 were the rapidly mutating Y-STRs (Romanos and Borjac, 2018). The mutation rate of all of the rapidly mutating Y-STRs was over 1×10^{-2} (Romanos & Borjac, 2018). The samples were extracted, quantified, amplified, and genotyped on the ABI 3500 Genetic Analyzer to confirm monozygotsity (Romanos & Borjac, 2018). The samples were re-amplified with the Y-filer Plus kit and genotyped. After genotyping, it was found that the Y-Filer Plus kit was unable to differentiate any of the pairs of monozygotic twins. Even though the results were not significant, a larger sample size and the use of NGS could aid in discovering differences in rapidly mutating Y-STRs between monozygotic twins.

RNA profiling: microRNAs

RNA profiling with miRNAs has also been used to discriminate monozygotic twins. miRNAs are small noncoding RNAs that help posttranscriptional gene regulation by binding to mRNAs (O'Brien et al., 2018). When miRNAs bind to mRNAs, they destabilize the mRNA and cause the protein not to be translated. As a result, miRNAs impact the level of expression a specific protein has without changing the gene sequence (O'Brien et al., 2018). miRNAs are small, stable, and easy to detect, which are all beneficial for forensic science. To determine if miRNAs are useful biomarkers for discriminating monozygotic twins, blood samples were collected from four pairs of monozygotic twins (Fang et al., 2019). Then the RNA was isolated, quantified, and sequenced on an Illumina HiSeq 2500. Once the miRNA profile was generated, it was found that 142–176 miRNAs were found in each of the samples (Fang et al., 2019). There was no significant difference found in the miRNAs present between the monozygotic twin pairs (Fang et al., 2019). As a result, the expression of the miRNAs was assessed for differences. In the first twin pair, 14 miRNAs were expressed at different levels (Fang et al., 2019). The second pair had 26 differently expressed miRNAs, the third pair had 10, and the fourth pair had 41. Only one miRNA, miR-451a, showed a significantly different level of expression in all of the twin pairs (Fang et al., 2019). Future research is required to determine if that miRNA is useful for the discrimination of monozygotic twins.

High-resolution melting to determine DNA methylation

Due to the fact that it is so rare and difficult to find differences in the nuclear and mitochondrial genomes, multiple researchers have turned to investigating the epigenome. The epigenome involves changes in DNA that occur because of environmental

events. One layer of the epigenome is DNA methylation. DNA methylation is the result of DNA methyltransferase adding a methyl group to the C5 position of cytosine. The methylation regulates gene expression and can be found in multiple locations, including CpG islands in the promoter region. To detect DNA methylation, the extracted DNA is treated with bisulfite in a reaction. Once treated with bisulfite, the unmethylated cytosine is converted into uracil, which is then converted into thymine during PCR. Methylated cytosine does not undergo bisulfite conversion and remains cytosine. There are two methods to detect the change from cytosine to thymine. One method is post-PCR high-resolution melt (PCR HRM). Researchers have observed differences in methylation patterns between monozygotic twins in Alu-E2F3 and Alu-SP (Stewart et al., 2015). These regions were bisulfite converted, and the melting temperatures were compared between the buccal swabs from five sets of twins. The more cytosine or higher methylation, the higher the melting temperature because cytosine-guanine is bonded with three Watson-Crick hydrogen bonds and thymine-adenine is held together by two Watson-Crick hydrogen bonds. The melting temperatures can be compared, and if the melting temperatures are different, it is concluded that the monozygotic twins have different methylation patterns in Alu-E2F3 or Alu-SP. After comparing the melting temperatures in a recent study, it was found that all five sets of twins had different methylation patterns in Alu-E2F3, and four sets of twins could be differentiated with Alu-SP (Stewart et al., 2015). This method is less time-consuming and less costly than sequencing methods (Stewart et al., 2015). Although it is more cost-efficient, it has a lower resolution and requires a larger amount of DNA to achieve adequate results due to its lower multiplexing capacity.

Detecting DNA methylation with pyrosequencing

Another method for differentiating monozygotic twins is to locate differently methylated regions (DMRs) with pyrosequencing. Pyrosequencing is useful for sequencing methylation sites or CpG sites in monozygotic twins in specific regions of the genome. One study investigated the methylation in LINE-1 (long interspersed element 1) (Xu et al., 2015). LINE-1 is a repetitive DNA retrotransposon (Ardeljan et al., 2017). This repetitive element makes up 17% of the human genome (Xu et al., 2015). Since it is a large portion of the genome, the methylation level of LINE-1 is used to infer the global DNA methylation within an individual (Xu et al., 2015). In addition, LINE-1 is made up of repeat sequences that have low selective pressure and greater variation, making the locus a good candidate for differentiating monozygotic twins (Xu et al., 2015). To analyze if the LINE-1 methylation level is different between monozygotic twins, buccal and blood samples from 119 pairs of monozygotic twins and 57 dizygotic twins were collected (Xu et al., 2015).

After sample collection and extraction, the samples were bisulfite-treated, amplified, and pyrosequenced at the promoter region of LINE-1 (Xu et al., 2015). The methylation level of the promoter region of LINE-1 was compared in the monozygotic twins, and t-tests were performed to determine if there was a statistically significant difference in the methylation level (Xu et al., 2015). The scientists observed that 15 out of 119 pairs of monozygotic twins were differentiated, and 10 out of 57 dizygotic twins were differentiated (Xu et al., 2015). There was correlation found with age in only the buccal samples methylation level. In addition, the sample type and sex of the individual were also found to impact the methylation level (Xu et al., 2015). It was found that the methylation level of the buccal samples was significantly lower than the methylation level in blood samples (Xu et al., 2015). Additionally, females had lower mean methylation values in blood samples but no difference in buccal samples (Xu et al., 2015).

Genome-wide DNA methylation profiling

Another approach to differentiating monozygotic twins is with NGS and genome-wide DNA methylation profiling with Illumina's BeadChip arrays. The Illumina Human Methylation 450k BeadChip array sequences over 450,000 CpG sites and the Illumina Infinium Human Methylation27 Beadchip kit sequences over 27,000 CpG sites (Li et al., 2013). Multiple studies have assessed the methylation pattern in monozygotic twins and found the most promising CpG sites for distinguishing monozygotic twins. One study analyzed blood samples from 22 pairs of monozygotic twins and the Illumina Infinium Human Methylation27 BeadChip assay to sequence thousands of CpG sites (Li et al., 2013). After sequencing, 92 CpG sites showed significant differences in methylation between the monozygotic twins (Li et al., 2013). These CpG sites were spread across 20 different chromosomes and were found to be related to the function of peripheral blood leukocytes (Li et al., 2013). These 92 CpG sites allowed for all 22 pairs of monozygotic twins to be distinguished (Li et al., 2013).

Another study also involved screening the blood samples of monozygotic twins to determine the main DMRs. Four pairs of monozygotic twins were screened and sequenced. After sequencing, 1772–3766 differently methylated regions were identified in each pair of monozygotic twins (Du et al., 2015). To narrow down the differently methylated regions to achieve the best results for differentiating monozygotic twins, criteria were set. The criteria included that the DMRs had to be present in all four sets of monozygotic twins and reach a defined level of methylation difference (Du et al., 2015). The methylation difference was calculated by mapping the reads, scoring the methylation, normalizing all of the scores, calculating the MeDIP scores, and then calculating the fold change between the MeDIP score of the monozygotic twins (Du et al., 2015). The fold change had to be greater than 830, with the MeDIP score in

one twin being zero while the other was larger than 8.31 to fit the criteria for DMR (Du et al., 2015). After applying these criteria, 38 DMRs were selected (Du et al., 2015). All of the DMRs that fit the criteria were in CpG islands. The DMR-associated genes were found to be involved in cell growth, differentiation, proliferation, and development (Du et al., 2015). One gene was a zinc finger protein-related gene. Zinc finger proteins are important in early development, and so it was hypothesized that the DMRs associated with zinc finger proteins are useful for discriminating younger twins (Du et al., 2015). This study shows that all four pairs of monozygotic twins could be differentiated with 38 DMRs, but the sample size was small, so further investigation on the 38 DMRs is needed to confirm that they are capable of differentiating all types of monozygotic twins (Du et al., 2015).

In addition to the above studies, buccal swabs, cigarette butts, saliva, and used cigarette buttss have been tested by genome-wide DNA methylation analysis. In a study, blood, buccal swabs, cigarette butt, and saliva samples were provided by one pair of 52.5-year-old European female monozygotic twins (Vidaki et al., 2018). The buccal cell samples were processed, extracted, quantified, and used to generate a genome-wide DNA methylation profile with the Illumina Infinium Human Methylation 450k BeadChip array. Each site was given a beta value, where 0 was assigned for completely unmethylated at the CpG site and 1 was assigned for completely methylated (Vidaki et al., 2018). The data generated was normalized, and DMRs were eliminated if the probes were mapped to multiple locations and if the marker had a beta value higher than 0.85 or lower than 0.15(Vidaki et al., 2018). In addition, the difference between the beta values had to be greater than 0.4. After considering these conditions, 129 candidate DMRs were found in buccal cells (Vidaki et al., 2018). For assay development, 22 of the top DMRs (over 0.4 difference) were selected (Vidaki et al., 2018). These sites were sequenced to determine methylation in saliva and cigarette butts. The saliva and cigarette butts had lower or opposite methylation differences in some of the 22 DMRs when compared to the buccal cell DNA. Therefore, the twins could not be differentiated with those sample types (Vidaki et al., 2018). Even though a difference in methylation in the twins could not be obtained with the saliva and cigarette butt samples, a difference was still obtained from the DNA extracted from the buccal swabs (Vidaki et al., 2018). This study expanded on previous studies, which only found DMRs in one sample type. Different cell types perform different functions than other cell types. Therefore, different genes will be methylated because different genes need to be downregulated or turned off. Buccal swabs, cigarette butts, and saliva are commonly found in casework. Therefore, finding DMRss in these sample types would be useful so that case evidence can be investigated no matter the sample type.

In another study, blood samples were collected from 12 pairs of monozygotic twins (Park et al., 2017). After sequencing with the Illumina Human Methylation 450k array,

over 450,000 CpG sites were narrowed down to 6 CpG sites (Park et al., 2017). The sites were narrowed down to the CpG sites that have the largest differences in methylation among the 12 pairs of twins (Park et al., 2017). The six CpG sites were cg00211609 (body of FAM132A), cg26287080 (body of EXOC7), cg01558909 (promoter of HBM), cg21036194 (body of SNCAIP), cg01419577 (intergenic regions), and cg04620228 (body of UTP14A) (Park et al., 2017). These six CpG sites were shown to have a high frequency of differences in methylation status, but further research was needed on these specific six sites to determine if additional twin pairs also have a high frequency of difference in methylation.

Using the sites identified, new pyrosequencing assays were developed. In the Towson Human Remains Identification Laboratory (THRIL) lab, we have performed pyrosequencing with cg00211609 (body of FAM132A) using 20 monozygotic twin sets. The twin-set samples were obtained from the Coriell Institute for Medical Research, extracted using the phenol-chloroform method, and quantitated with the Qiagen Quantiplex Pro RGQ Kit using a Qiagen RotorGene Q real-time PCR instrument. After quantitation, the samples were bisulfite-treated with the Qiagen EpiTect Plus Bisulfite Conversion Kit. After bisulfite treatment, the samples were amplified with the Qiagen PyroMark PCR kit and sequenced on a Qiagen Q48 pyrosequencing instrument with the Qiagen Gene Globe assay, which includes the primer cg00211609. Out of the first 13 monozygotic twin pairs that were successfully sequenced, several pairs exhibited a statistically significant difference in percent methylation when using a paired t-test. Even though some twin pairs could not be differentiated with the one primer, other primers and CpG sites may display differences in methylation. We are continuing to examine additional loci, including cg26287080 (body of EXOC7) and cg01558909 (promoter of HBM), with more sets of monozygotic twins.

Conclusion

The advancements in NGS technologies have allowed for higher throughput, so that large portions of a suspect's genome can be sequenced. Sanger sequencing, NGS, pyrosequencing, and PCR HRM can be used to differentiate monozygotic twins. With this technology, differences in the mitochondrial genome, nuclear genome, miRNAs, and DNA methylation patterns can be detected. These differences can be used to differentiate monozygotic twins, which cannot be differentiated with STR typing due to monozygotic twins having identical STR profiles. Multiple studies have been released finding differences in monozygotic twins, with the various research studies focusing on different portions of the nuclear or mitochondrial genome. In addition, with the exception of high-resolution melt, the methods discussed are expensive to perform with expensive instrumentation not widely available in forensic laboratories, which can be an issue for their implementation in forensic criminal casework. Tissue variation remains a challenge

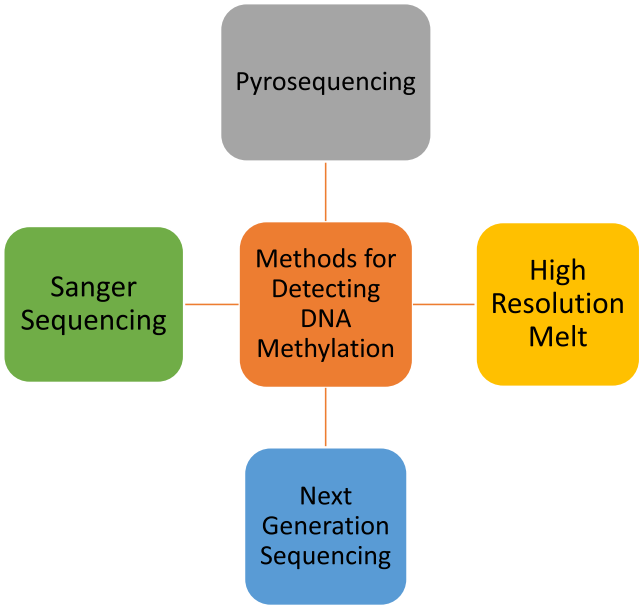


Figure 10.2 Methods for detecting DNA methylation.

for identifying differences in monozygotic twins that vary across all sample types. Therefore, further research is required to identify loci that can be used to differentiate the most pairs of monozygotic twins, decrease the cost, and standardize and validate the methods so that one or more can be used in criminal cases and be accepted into forensics (Fig. 10.2).

Acknowledgments

We would like to thank the Kaela Green for her presentation on Forensic DNA Analysis of Identical Twins at Towson University and Mary Jones-Dukes, Frank Fischinger, Adam Klavens, and Sonia Vagues for their technical assistance with the twin study.

References

Ardeljan, D., Taylor, M., Ting, D., & Burns, K. (2017). The human LINE-1 retrotransposon: an emerging biomarker of neoplasia. *Clincal Chemistry, 63*(4), 816–822. https://doi.org/10.1373/clinchem.2016.257444

Du, Q., Zhu, G., Fu, G., Zhang, X., Fu, L., Li, S., & Cong, B. (2015). A genome-wide scan of DNA methylation markers for distinguishing monozygotic twins. *Twin Research and Human Genetics, 18*(6), 670–679. https://doi.org/10.1017/thg.2015.73

Fang, C., Zhao, J., Liu, X., Zhang, J., Cao, Y., Yang, Y., Chunrui, Y., Xiaoli, Z., Jialin, Q., Wenli, L., Huijuan, W., & Jiangwei, Y. (2019). MicroRNA profile analysis for discrimination of monozygotic twins using massively parallel sequencing and real-time PCR. *Forensic Science International: Genetics, 38*, 23–31. https://doi.org/10.1016/j.fsigen.2018.09.011

Li, C., Zhao, S., Zhang, N., Zhang, S., & Hou, Y. (2013). Differences of DNA methylation profiles between monozygotic twins' blood samples. *Molecular Biology Reports, 40*(9), 5275–5280. https://doi.org/10.1007/s11033-013-2627-y

O'Brien, J., Hayder, H., Zayed, Y., & Peng, C. (2018). Overview of MicroRNA Biogenesis, Mechanisms of Actions, and Circulation. *Frontiers in Endocrinology, 9*. https://doi.org/10.3389/fendo.2018.00402

Park, J.-L., Woo, K.-M., Kim, S.-Y., & Kim, Y. S. (2017). Potential forensic application of DNA methylation to identify individuals in a pair of monozygotic twins. *Forensic Science International: Genetics Supplement Series, 6*. https://doi.org/10.1016/j.fsigss.2017.09.177

Romanos, P., & Borjac, J. (2018). Could rapidly mutating Y-STRs be a potential forensic tool in discriminating Lebanese monozygotic twins? *Genetics & Applications, 2*(2), 35–42. https://doi.org/10.31383/ga.vol2iss2pp35-42

Stewart, L., Evans, N., Bexon, K. J., van der Meer, D. J., & Williams, G. A. (2015). Differentiating between monozygotic twins through DNA methylation-specific high-resolution melt curve analysis. *Analytical Biochemistry, 476*, 36–39. https://doi.org/10.1016/j.ab.2015.02.001

Vidaki, A., Kalamara, V., Carnero-Montoro, E., Spector, T., Bell, J., & Kayser, M. (2018). Investigating the epigenetic discrimination of identical twins using buccal swabs, saliva, and cigarette butts in the forensic setting. *Genes, 9*(5), 252. https://doi.org/10.3390/genes9050252

Wang, Z., Zhang, S., Bian, Y., & Li, C. (2015). Differentiating between monozygotic twins in forensic through next-generation mtGenome sequencing. *Forensic Science International: Genetics Supplement Series, 5*. https://doi.org/10.1016/j.fsigss.2015.09.023

Weber-Lehmann, J., Schilling, E., Gradl, G., Richter, D. C., Wiehler, J., & Rolf, B. (2014). Finding the needle in the haystack: Differentiating "identical" twins in Paternity Testing and forensics by ultra-deep next generation sequencing. *Forensic Science International: Genetics, 9*, 42–46. https://doi.org/10.1016/j.fsigen.2013.10.015

Xu, J., Fu, G., Yan, L., Craig, J. M., Zhang, X., Fu, L., Ma, C., & Li, S. (2015). Line-1 DNA methylation: A potential forensic marker for discriminating monozygotic twins. *Forensic Science International: Genetics, 19*, 136–145. https://doi.org/10.1016/j.fsigen.2015.07.014

Yuan, L., Chen, X., Liu, Z., Liu, Q., Song, A., Bao, G., Wei, G., Zhang, S., Lu, J., & Wu, Y. (2020). Identification of the perpetrator among identical twins using next-generation sequencing technology: A case report. *Forensic Science International: Genetics, 44*, 102167. https://doi.org/10.1016/j.fsigen.2019.102167

CHAPTER 11

Application of NGS in maternal genome analysis in ancient human remains

Muhammad Irfan, Khushbukhat Khan and Naila Khan
Department of Healthcare Biotechnology, National University of Science and Technology Islamabad, Atta-ur-Rahman School of Applied Biosciences, Islamabad, Pakistan

Introduction

An individual's genetic uniqueness is determined by the combination of genetic markers inherited from his or her parents. A genetic marker is a gene or a short segment of DNA that has a distinctive location on a specific chromosome that can be used for the identification of a species or an organism. It can be as short as a single nucleotide polymorphism or a longer sequence, such as mini and microsatellites. Genes on a chromosome exist in alternate forms known as alleles that differ in sequence. The goal of genetic marker typing is to determine which alleles are present at genetically variable loci (Jordan & Mills, 2021; Khan et al., 2017, 2020). There are three types of DNA markers: Autosomal, paternal, and maternal. Autosomal markers are unique DNA sequences scattered across the genome that are used in forensics, population genetics, pedigree analysis, missing persons' identification, and genealogy. Recently, the interest of scientists has shifted toward the use of maternal or mitochondrial DNA (mtDNA) markers for genetic analysis such as the identity of missing persons, linkage of suspects to the crime scene, and all sorts of evolutionary and genealogical analysis. The use of mtDNA for the identification of an organism especially from ancient remains has several advantages over autosomal DNA. First, the presence of a high copy number of mtDNA in each cell as compared to autosomal DNA, which has only two copies per diploid cell (Cole, 2016; Michael Hofreiter et al., 2021; Parson et al., 2013). Second, mtDNA is maternally inherited and follows the nonmendelian inheritance pattern, thus lacking recombination, making it easier to trace unique sequences through the generations (Zhang et al., 2018). Moreover, in ancient remains the DNA is often found degraded while the mtDNA is more likely to survive for longer periods, providing enough DNA to be sequenced and compared. These advantages make mtDNA a powerful tool to be used in evolutionary and genealogical studies extensively applied to ancient samples, tying ancient remains with living descendants (Chyleński et al., 2019; Cui et al., 2013; Hofreiter et al., 2021; Modi et al., 2017; Willerslev & Cooper, 2005).

Next Generation Sequencing (NGS) Technology in DNA Analysis
ISBN 978-0-323-99144-5, https://doi.org/10.1016/B978-0-323-99144-5.00005-6
© 2024 Elsevier Inc.
All rights reserved.

Due to the practical and technical hurdles involved in the sequencing of the entire mitochondrial genome using the Sanger sequencing method, the analysis of mtDNA is limited to a few nucleotide-long segments approximately 700 bases), i.e., hypervariable regions I and II (HVI and HVII). Most of the technical difficulties are addressed by the introduction of new high-throughput technology, including next-generation sequencing) (NGS) methods. The first of these NGS approaches was reported in 2005 and implemented in ancient DNA (aDNA) research (Margulies et al., 2005). These methods have enabled the sequencing of millions of fragments at once, reducing the cost and complexity associated with the sequencing of ancient mtDNA. The use of such high-throughput methods for whole mitochondrial genome analysis resulted in an increase in the percentage of unique haplotypes from 65% to 90%–100%. In addition, unlike conventional sequencing techniques, NGS allows for the detection of misincorporated patterns, which aids in distinguishing between endogenous aDNA and current contaminants (Kirkinen et al., 2022; Senovska et al., 2021).

This chapter discusses the key role NGS has played in unraveling the historical secrets and evolutionary patterns by generating a bulk of data based on maternal DNA signatures. The chapter also discusses the procedure opted to prepare aDNA samples for NGS, mtDNA capture from highly degraded samples, and bioinformatics tools used for the analysis. The chapter also sheds light on the comparative advantage of using NGS over traditional techniques for mtDNA profiling from ancient samples.

What stories ancient DNA analysis tell us?

aDNA analysis helps in addressing a number of important questions that were previously impossible to answer. The majority of aDNA research is focused on the origin, relationship, genetic diversity, and migration of human populations. The classification of human remains into a specific group or a well-defined population is performed using aDNA analysis (Carlyle et al., 2000; Keyser-Tracqui et al., 2003; Kolman, 1999, pp. 183–200; Krings et al., 2000; Lalueza et al., 1997). The aDNA retrieved from the remains of hominins and Neanderthals has been analyzed to clarify their relationships with present-day humans (Gibbons, 2021; Hawks, 2021; Vernot et al., 2021).

aDNA analysis has enabled scientists to get insight into the reasons behind extinction by analysis of extinct species. The DNA from the ancient remains of different extinct creatures has been extracted and analyzed to investigate the genetic variations. Obtained outcomes are compared to the data from genomic libraries containing information regarding the genomic signatures of different species from the past (if studied) and present. Then attempts are made to find the possible cause behind the widespread extinction of these species (Brüniche–Olsen et al., 2018; Orlando et al., 2002; Woods et al., 2021).

The analysis of aDNA has been used for the identification of the sex of ancient human skeletons using genetic markers on the X and Y chromosomes. In most cases,

the remains found are degraded, due to which the sex cannot be determined morphologically. The data obtained from such studies can be used to shed light on marriage patterns, burial customs, and differential rates of mortality (Nistelberger et al., 2019; Wasef et al., 2020).

The identification of kin relationships in the case of group burial is also one of the applications of aDNA analysis. When a group of human skeletons is discovered, it is of interest to find out if the individuals are related to a family unit. For example, a study involving 62 specimens from a 2000 BP necropolis was conducted to reconstruct the kinship of specimens. The DNA from skeletal remains was extracted and amplified for single-copy autosomal microsatellite markers. The results revealed close relationships, including few parent/child relations (Brück, 2021; Drosou et al., 2018; Gad et al., 2021; Keyser-Tracqui et al., 2003).

The genetic analysis of ancient fecal samples can be used to investigate dietary patterns and find out what plants and animals are consumed by the defecating individual. Moreover, the fecal samples can also be used as an alternate source of DNA to identify the individuals, as in some cases the coprolites provide a better source of aDNA as compared to skeletal or mummified tissues (Hofreiter et al., 2000; Poinar et al., 1998).

Nonhuman remains are also subjected to aDNA analysis to investigate certain aspects of human prehistory and to reconstruct and enhance our understanding of the domestication process and agriculture. The DNA from animal remains can be extracted using the same techniques used for human DNA (Barnes et al., 2002; Jaenicke-Després et al., 2003; Leonard et al., 2002). The aDNA analysis has shown that present-day dogs are known to have interbreed with dogs brought by European colonists (Leonard et al., 2002). Similarly, analysis of an ancient maize sample from Mexico showed that alleles in each of the three important genes that are predominant in maize nowadays were already present 4000 years ago (Jaenicke-Després et al., 2003). The analysis of ancient floral and faunal remains helps us to understand the ecosystem in which prehistoric humans lived; furthermore, the analysis of the molecular diversity of animal and plant remains also gives clues about the hunting and dietary habits of humans. The analysis of ancient animal remains also attempted to explain the movements of prehistoric populations (Matisoo-Smith et al., 1997).

aDNA analysis addresses another essential question related to the evolution of human pathogens and the pattern of prehistoric diseases (Gad et al., 2021). aDNA studies have been performed to identify different human pathogens such as anthrax, bubonic plaque, tuberculosis (TB), and influenza (Basler et al., 2001; Donoghue et al., 1998; Drancourt & Raoult, 2002). The spread of TB, caused by *Mycobacterium tuberculosis*, was traced back to European colonization; however, the aDNA analysis identified TB in a 1400-year-old Byzantine fragment of calcified lung tissue, which confirmed the presence of the pathogen before European contact (Donoghue et al., 1998, 2004).

Challenges faced in studying ancient DNA

Investigations based on aDNA have allowed scientists to understand the major historical events that shaped the fate of modern humans. However, the road to this analysis was full of challenges. Despite huge advancements in molecular biology techniques, some of these challenges are still faced by researchers.

Contamination

Amplification of contaminant DNA is a problem in all studies involving polymerase chain reaction (PCR); however, this problem is even more severe when dealing with ancient human DNA, which is often found in very low amounts. The ancient and historic biological samples discovered at excavation sites yield a very low amount of DNA; moreover, this DNA is often contaminated and modified in different ways, which makes it difficult to amplify the authentic DNA extracted from these samples. Contamination of a single molecule of modern DNA can cause errors in the analysis (Furtwängler et al., 2018; Grigorenko et al., 2009). Even with the use of highly specific and rigorous procedures (Hofreiter et al., 2001; Poinar & Cooper, 2000), modern human DNA contamination is still very prevalent in the amplification of aDNA extracted from ancient remains (Hofreiter, n.d.; Krings et al., 2000). In the case of bone and teeth ancient samples, it is almost impossible to completely purify the endogenous DNA from modern DNA contamination. Decontamination procedures such as ultraviolet C (UVC) and bleach treatments are also not efficient enough to fully eradicate the contamination (Gilbert et al., 2005). That may be due to the presence of pores in the bone and tooth dentine that serve as the main entry sites for DNA, especially from sweat, skin fragments, and exhaled cells. The teeth are less prone to contamination when present in well-preserved form as compared to bone fragments. Hairs also seem to be less prone to contamination and are considered one of the most reliable sources for ancient human DNA analysis, despite their less common presence in ancient samples as compared to skeletal remains (Gilbert et al., 2004).

Different studies have reported the presence of modern human DNA contamination in ancient samples. For example, a study on Neolithic and Upper Palaeolithic samples discovered in Spain sequenced the mtDNA hypervariable region 1 (HVR1) and compared it with the DNA obtained from six individuals working on these samples. The comparison of their sequences showed that 17.13% of the 572 cloned sequences from aDNA were actually contaminants from the individuals working on the project (Caramelli et al., 2008; Sampietro et al., 2005).

Chemical modifications and degradation of ancient DNA

The alterations and mutations in postmortem DNA are also one of the main problems in complex aDNA analysis. The amplification of aDNA becomes very difficult compared to DNA extracted from a fresh sample because of the alterations and degradations of aDNA.

One of the most obvious types of DNA damage of aDNA is its degradation to small fragments. This reduction in the size of DNA fragments is caused by enzymatic actions that occur right after the death and also nonenzymatic hydrolytic cleavage of phosphodiester bonds creating single-stranded nicks (Geigl & Grange, 2018; Lindahl, 1993; Shapiro, 1981, pp. 3–18). The hydrolytic cleavage of glycosidic bonds between nitrogenous bases and the sugar backbone results in abasic sites, which then undergo chemical rearrangement leading to strand breakage (Mohni et al., 2019; Lindahl & Nyberg, 1972; Lindahl & Karlstrom, 1973; Shapiro, 1981, pp. 3–18). The extent of DNA degradation depends on the sample age, the geographic origin of the sample, and also the environmental conditions where the sample was stored (Grigorenko et al., 2009; Gabriel Renaud et al., 2019).

Unlike metabolically active cells, postmortem cells that are in dormant conditions do not have an active DNA repair system and are prone to hydrolytic or oxidative DNA modifications and strand damage. The most critical modification the DNA undergoes is the one that induces alteration in the bases but does not halt the process of amplification, leading to nucleotide modification in the amplification products. These modifications include type I substitutions (A > G/T > C) and type II substitutions (C > T/G > A). Different methods are used in aDNA analysis to minimize postmortem DNA modification, which is a prerequisite to enhance the overall quality and concentration of DNA templates. These methods make use of two important enzymes: Uracil-*N*-glycosylase for eliminating the deamination products of cytosine (Gilbert et al., 2003; Hofreiter et al., 2001) and *N*-phenacylthiazolium bromide necessary for breaking the intermolecular cross-links (Mohni et al., 2019). Moreover, the application of high-fidelity DNA polymerase such as pfu *Taq HiFi* also reduces the rate of sequence errors, thus increasing the efficacy of aDNA amplification (Cooper et al., 2001).

Next-generation sequencing for maternal genome analysis in ancient samples

NGS has provided solutions to several limitations that were halting the scientists from gathering the information contained in human ancient samples. One of the key limitations faced in the analysis was contamination from several sources, which challenged the authenticity of the obtained results. In 2006, two studies were performed on the same Neanderthal sample, and the outcomes of the studies produced inconsistent results due to sample contamination (Green et al., 2006; Noonan et al., 2006). As a consequence of the faced issues, a set of criteria has been set to reduce the risk of sample contamination. A key step in eliminating the maximum possibility of contamination is the cleanliness of the NGS facility (Green et al., 2006, 2009; Wall & Kim, 2007). Experimentation revealed that samples prepared for NGS in a cleaned environment had minimal contamination of the modern human genome. Contrarily, sample preparation commenced in a nonclean laboratory had modern human DNA as a contaminant (Green et al., 2009).

Despite the ancient sample contamination with the modern human genome, a new set of tools in NGS provides benefits in analyzing aDNA by identifying and avoiding contamination. aDNA has genomic material with very short sequences. NGS, in comparison to PCR, allows the detection and sequencing of those molecules and avoids the amplification of larger genomic segments (Krause., et al., 2010).

Ancient samples are sensitive, and special care is required for their preparation for NGS. Similarly, sample contamination with other sources, such as human, bacterial, or fungal, often compromises the result of aDNA experimentation (Knapp et al., 2012; Miller et al., 2009). Therefore, caution is practiced when preparing samples from ancient sources to get maximum results.

Facility for ancient sample processing

The spatial separation of the aDNA facility from the postPCR processing laboratory is essential to avoid the chances of cross-contamination. Further, the movement from the aDNA facility to the postPCR laboratory should be unidirectional, and laboratory workers must establish a practice to visit aDNA facility before visiting the postPCR room. Multiple visits to the aDNA facility from a postPCR laboratory on the same day enhance the risk of contamination. Therefore, a facility should be designed in such a way that entrance to aDNA facility should not be from a postPCR laboratory (Farrer et al., 2021; Knapp et al., 2012; Pinhasi et al., 2019). Further, such facilities often have a limited-access policy, and access to aDNA facility is only available to trained individuals who are aware of the protocols necessary for reducing the contamination risk.

aDNA facility spatial isolation from postPCR laboratory must be accompanied by the separation of space for conducting the different steps of aDNA processing in the aDNA facility. As steps for processing ancient samples can vary because of the nature of the sample, therefore, it is acceptable to at least have three separate rooms or hoods for conducting specific experimental activities (Fulton, 2012). One is for changing personal protective equipment, consumables storage, UVC irradiation-based decontamination of samples and consumables, and preprocessing of museum samples that may be contaminated. The second room is for the processing (cutting or grinding) of bones. This room should ideally have an internal UVC-equipped fume hood. The third room is for performing DNA extraction and DNA amplification experiments (Templeton et al., 2013).

Protocols for reducing contamination are particularly necessary for preparing aDNA samples for NGS. A little amount of nontarget DNA can be a source of nuisance, interfering with the sequence read of sample DNA. Thus, reducing the cross-contamination risk using hairnets, gloves, facemasks/shields, HEPA-filtered air conditioning, positive-pressure systems, fume hoods, and laminar flow hoods is a prerequisite for achieving nuisance-free peaks in NGS sequence plots.

Sample preparation and DNA extraction from ancient samples

Samples for NGS are prepared in an ultraclean environment to avoid modern DNA contamination, and much deliberation is practiced in dealing with ancient samples. Unlike fresh specimens, ancient samples lack cells or tissues, a possible reason for the low-available concentrations of endogenous DNA. Similarly, ancient samples are often going through various states of degradation or are severely damaged. Therefore, the methodology for the extraction of DNA must not be so rigorous, and the use of chemical detergents and high temperatures must be avoided. Although such procedures result in a high release of DNA, they could inflict more damage to the aDNA, leading to a poor overall yield of DNA. Lastly, ancient samples are contaminated with numerous PCR inhibitors that may interfere with the DNA amplification processes. So, the method for DNA extraction from ancient bone or tooth samples must be efficient enough to deal with these problems. Two methods used for the initial processing of ancient samples are the Loreille and Dabney protocols. These protocols are shown to work best for ancient samples and burned samples, respectively. However, alteration in these protocols based on the amount of initial powdered sample used and the composition of lysis buffer is shown to make these methods more effective for high-throughout the mtDNA analysis in ancient samples (Xavier et al., 2021).

These methods to obtain DNA from ancient samples are destructive for the sample itself and majorly rely on drilling, powdering, or cutting. On the obtained powdered samples, methods such as isopropanol or ethanol precipitation, isolation through membranes, or silica method can be employed for DNA molecule extraction (Damgaard et al., 2015; Rohland & Hofreiter, 2007). Silica method to date is the most preferred method due to the comparatively high yield of DNA (Gamba et al., 2014; Rohland et al., 2010, 2018). This is a 2-day method that begins with the removal of surface dirt that may contain several sorts of inhibitors. Such inhibitors can interfere with the extraction process or disrupt the enzymatic manipulation of the extracted DNA at later stages. The outer surface of the ancient sample is carefully removed using a single-use grinding tool. This step is critical as it can enhance the endogenous-to-exogenous DNA ratio. In the next step, a compact part of the bone is cut off, or a tooth root or dentine is removed. As such procedures are irreversibly damaging, curators usually do not allow the removal of parts belonging to unique specimens. In such a scenario, a drilling techniques are employed, and fine powder is obtained without major damage to the sample. Drilling is performed at low speed to avoid heat generation that could possibly deteriorate aDNA (Dabney et al., 2013; Rohland & Hofreiter, 2007). In a few labs, tooth samples are cavitated and ultrasonically cleaned prior to drilling, while the UV-irradiation method is employed in other laboratories to decontaminate samples (Harney et al., 2021; Matsvay et al., 2019). The finely ground powder is obtained by mortar and pestle crushing, a ball

mill, or a freezer mill (Matsvay et al., 2019; Rohland & Hofreiter, 2007). The finer the powder; the more the DNA yield. The powdered sample is then mixed with the extraction solution and incubated overnight on gentle shaking in the dark (Rohland & Hofreiter, 2007). One study suggested that predigestion of powdered samples with EDTA (0.5M), proteinase K, and N-laurylsarcosyl solution improves the yield by ~ 2.5-fold (Damgaard et al., 2015). The next day, a sample is mixed with binding buffer and silica suspension, and pH is maintained between 4.0 and 3.5. pH below this range may damage DNA. The solution is incubated for 3 h at agitation for allowing DNA binding with silica. The choice of binding buffer type and purification matrix greatly impacts the yield and DNA fragment that will be eluted in later stages (Rohland et al., 2018). Once a DNA-silica suspension is obtained, binding buffer is removed through washing, and DNA is eluted through DNA elution buffer (Rohland & Hofreiter, 2007).

Recent investigations reported the methods to minimize ancient sample destruction during its processing. The study demonstrated the use of targeted bleaching at the ancient tooth root. Most studies preferred the tooth root or dentin for extraction purposes. A study by Harney and his team extracted DNA from the cementum of the tooth. A bone-like layer, cementum, is present in the root of the tooth that is enriched with cementocytes. These cells are preserved after death under the mineral structures of the tooth and are a good source of DNA. Harney and colleagues exposed ancient tooth roots to lysis solution for a minimal duration, which caused less destruction to the sample, and the yield of DNA was also equivalent to the destructive processing methods for ancient samples (Harney et al., 2021). Despite the availability of numerous extraction protocols, it will be wise to choose an extraction method based on sample type, degree of degradation, and sample age.

Commercially available kits are also employed for extracting DNA from ancient samples, which more or less work on the same principles. Recently, the commercially available kit PrepFiler BTA forensic DNA extraction kit was used to extract aDNA from human tooth samples. This kit uses magnetic beads for DNA extraction. This kit is designed for extracting degraded DNA from forensic evidence and provides its application in dealing with ancient samples (Matsvay et al., 2019). The concentration of extracted aDNA is usually determined through a Qubit dsDNA HS Assay Kit on a Qubit 2.0 fluorimeter (Matsvay et al., 2019; Neparáczki et al., 2017).

A special module system was introduced in 2018 to process ancient samples for NGS. The sample consisted of teeth from the skull discovered during the seventeenth-century burial site excavation in Radonezh, Russia. Both sample preparation and DNA extraction were performed in the four-glove box module system. Antechambers connected each box and had an evacuation system that allowed a high purity atmosphere for the preparation of samples. Molecular filters, present in the air purification system, prevented the entrance of molecules larger than nitrogen atoms maintained the sterility of the boxes, and allowed sample processing in the oxygen and organic molecules' absence. Each step

of the extraction was conducted in a separate chamber. This system was beneficial as each box was connected, and the chances of contamination during sample transfer from one lab to another were also avoided (Matsvay et al., 2019).

Library preparation for next-generation sequencing

aDNA is often degraded, and its segments have blunted or overhanging ends. Additionally, aDNA can possess deaminated cytosines, converting them into uracil bases. aDNAs are repaired for these shortcomings using two different methods. Through the first method, uracil bases from the sequence are removed, and flanking 5′ and 3′ ends are repaired to produce a ligatable blunt end. The reaction is performed by adding three enzymes: T4 DNA Polymerase (Thermo Fisher Scientific Inc.), T4 PolyNucleotide Kinase (T4 PNK, Thermo Fisher Scientific Inc.), and T4 DNA Ligase (Thermo Fisher Scientific Inc.). The first two enzymes repair the overhanging ends through their 3′-5′ exonuclease activity and phosphorylation ability, respectively. The last enzyme ligates the blunt ends to synthetic adapters. Other methods only remove flanking regions and do not excise uracil (Briggs & Heyn, 2012; Matsvay et al., 2019). Repaired aDNA is then purified using spin columns such as the MinElute spin column, QIAGEN, or carboxyl-coated magnetic beads (Sera-Mag SpeedBeads) (Briggs & Heyn, 2012; Neparáczki et al., 2017). In the next step, adapter sequences are attached to the repaired sequences, and the entire library is subjected to amplification through PCR. The amplification procedure makes use of the universal priming sites that are introduced through adapters. The final concentration for the amplified product should be 50 ng/μL and the libraries below must be reamplified (Neparáczki et al., 2017). The use of Phusion enzyme a proofreading enzyme is recommended during second amplification to avoid PCR errors (Briggs & Heyn, 2012).

Mitochondrial DNA capturing and sequencing

mtDNA-capturing biotinylated probes are prepared from long-range, overlapping amplified products. These probes are designed using NimbleGen, a proprietary software developed by Calloway and his team (Calloway et al., 2015). This software employs the information from a human mitochondrial genome population database, global mtDB, and targets the consensus sequences in mtDNA. The tiling approach is applied in probe designing to obtain targeted overlapping sequences with high redundancy. These probes are 50−100 bp in length that could effectively target most of the mtDNA genome (Shih et al., 2018). Roche NimbleGen SeqCap EZ Library Developer Library kit is generally used during hybridization experiments. Samples for hybridization consist of probes (100 ng) and library DNA (400 ng). The thermal profile generally followed for hybridization is DNA denaturation for 5 min at 95°C and overnight (14−18 h) incubation to allow probes to bind to complementary sequences on library DNA. Blocking oligonucleotides are also included in the mix to shun nonspecific hybridization and promote strand displacement (Neparáczki et al., 2017; Shih et al., 2018; Templeton et al.,

2013). The number of probes varies from two to four. The sequence of probes generally used is below: BO1.P5.part1F: AATGATACGGCGACCACCGAGATCTACAC-Phosphate; BO2.P5.part2F ACACTCTTTCCCTACACGACGCTCTTCCGATCT-Phosphate; BO4.P7.part1 R GTGACTGGAGTTCAGACGTGTGCTCTTCCGA TCT-Phosphate; and BO6.P7.part2 R CAAGCAGAAGACGGCATACGAGAT-Phosphate. Dynabeads MyOne Streptavidin C1 magnetic beads (Thermo Fisher) are used to capture biotinylated probe-DNA complexes, and release of library DNA is achieved by treating the captured complex with the strand-displacing Bst DNA polymerase enzyme (large fragment, New England Biolabs). Amplification of hybridized products is accomplished through KAPA HiFi HotStart Ready Mix (KAPA Biosystems) and amplification is done for 10−15 PCR cycles. The amplified product is again purified using a MinElute column or magnetic beads, the purified product is quantified and subjected to sequencing. Systems used for sequencing are the MiSeq sequencer, and the sequencing reaction is prepared using MiSeq Reagent Kit v3 (Illumina, MS-102−3003) (Shih et al., 2018; Templeton et al., 2013), Ion Torrent Personal Genome Machine (PGM; Life Technologies), and Ion PGM 200 sequencing kit v2 chemistry (Life Technologies) (Templeton et al., 2013).

Next-generation sequencing data analysis

Software used for studying the mtDNA data obtained after NGS are NextGENe v.2.4.2 (SoftGenetics LLC., State College, PA, USA) and GeneMarkerHTS v.1.2.2. Resulting reads are mapped on the genome assembly using Burrows Wheeler Aligner v0.7.9 software. Initially, the genome assembly GRCh37.75 human genome reference sequence was used to access the revised Cambridge Reference Sequence (rCRS, NC_012,920.1) for mtDNA (Mudge et al., 2021; Neparáczki et al., 2017). Recent studies, however, prefer the Reconstructed Sapiens Reference Sequence for mapping mtDNA haplotype data (Behar et al., 2012; Sarmento, n.d.). More tools such as Picard Tools v 1.113 for data cleaning for duplicate occurrence, SAMtools v1.1 for data sorting and indexing (Etherington et al., 2015), mapDamage2.0 for aDNA damaged pattern assessment (Jónsson et al., 2013), and FreeBayes v1.02 for genetic variant identification (Garrison & Marth, 2012).

Next-generation sequencing based on ancient sample mtDNA in evolutionary studies

DNA samples from ancient remains were largely studied using PCR and Sanger sequencing before the invention of high-throughput NGS. Sanger sequencing was once thought to be the gold standard for DNA sequencing; however, it had a low throughput and was hence expensive, especially for large-scale sequencing investigations.

Furthermore, library preparation and amplification, as well as colony selection, are time-consuming and inefficient processes (Rizzi et al., 2012). These limitations of traditional sequencing are even more critical in the case of aDNA, and the emergence of NGS has opened up entirely new possibilities and expanded the area of applications (Millar et al., 2008). The new generation of NGS sequencers has transformed aDNA research by increasing the number of bases sequenced per run while decreasing sequencing costs. The most widely used NGS technologies in the field of aDNA analysis are the 454/Roche FLX and the Illumina Genome Analyzer (Margulies et al., 2005; Millar et al., 2008; Shendure et al., 2004). A significant amount of research has been done utilizing the applications of NGS in analyzing mtDNA retrieved from ancient samples.

Mitochondrial DNA from coat hair of mammoth

Hair shafts are a promising source of aDNA. Several properties of shafts suggest that they constitute an attractive DNA source for NGS. They are mostly present in abundance, which makes them preferable to bones, as the destructive sampling can cause the loss of important morphological information. Moreover, the turnover of keratinocytes in the hair bulb is significantly high; therefore, the mitochondrial levels in these cells may be comparatively higher than those in the other cells used for the analysis of aDNA (Vanscott et al., 1963). mtDNA was extracted from the coat-hair shafts of a Siberian woolly mammoth *(Mammuthus primigenius)* found in northern Siberia's permafrost deposits. The use of hair shafts in combination with sequence by synthesis resulted in 10 full mitochondrial genome sequences with 7.3−48.0-fold coverage. The yield of mtDNA sequence was about 26 times higher than from the previously reported permafrost-preserved bone (Poinar et al., 2006). Even though one of the specimens was $>$ 50,000 14C years old and another had been maintained at ambient temperature for 200 years, the observed levels of sequencing errors were lower than those found in previously published frozen bone samples. This method therefore prepared the groundwork for molecular-genetic examination of museum collections. Furthermore, the rate of damage-derived sequencing errors was also reported to be lower than those observed in previously published frozen bone samples, irrespective of the age of the sample, as one of the specimens included in the study was $>$ 50,000 years old (Gilbert et al., 2007).

After that, five new full genomes of Siberian woolly mammoths were sequenced using 454 sequencing NGS technology coupled with conventional PCR-based approaches with up to 73-fold coverage from the hair shaft. Three out of these five genomes represent the first-ever clade II mammoth genomes. The complete analysis of these and previously reported mitochondrial genomes demonstrated that there are two mitochondrial clades (clade I and clade II), both of which are highly diverged. Furthermore, statistical examination of the combined datasets revealed that clade II disappeared from Siberia long before clade I (Gilbert et al., 2008).

Mitochondrial DNA of Paleo-Eskimo

The Independence I-Saqqaq and Pre-Dorset Cultures, which covered roughly 3900–2500 14C years before now (BP), and the Independence II/Dorset Cultures, which spanned around 2500–700 14C years before present (BP), are known collectively as the Paleo-Eskimos (Eskimo, 1984). They represent the earliest human expansion into the northernmost reaches of the New World. The classification of modern and ancient Neo-Eskimo has been performed using the analysis of the hypervariable I (HVS1) region of mtDNA. A 3400–4500-year-old frozen hair sample was retrieved from an early Saqqaq settlement of paleo-Eskimo humans and sequenced using NGS. The sequencing data revealed that the sample varies from modern Native Americans and Neo-Eskimos and belongs to haplogroup D2a1 similar to the one previously found in modern Aleuts and Siberian Sireniki Yuits. These findings show that the migrations that brought early people to the far north of the New World did not come from Native Americans or the groups who eventually gave rise to the Neo-Eskimo expansion from Alaska around 1000 years ago (Gilbert et al., 2008).

The genetic history of Neanderthals (the divergence between extant humans and Neanderthals)

Neanderthals were a type of hominid that lived in Europe and Western Asia until approximately 30,000 years ago. They are believed to be more closely related to modern humans than chimps, which is why studying their genomes provides a unique opportunity to discover genetic modifications that are specific to anatomically fully developed modern humans. The first molecular genetic analysis of Neanderthals was performed in 1997. A Neandertal-type specimen was discovered in the Neander Valley, near Düsseldorf, Germany, in 1856. A 379-base-pair-long section of HVRI of the mitochondrial genome was determined (Krings et al., 1997).

High-throughput 454 sequencing techniques have recently been used to study the mtDNA of Neanderthals. The main advantage of using 454 sequencing technologies is the huge volume of sequence data that is generated by these technologies, making aDNA sequencing studies feasible. These NGS technologies also help in increasing our understandings of the modifications and degradation of DNA that occur during deposition in burial contexts.

One of the studies reported an entire mitochondrial genome from the bone of a 38,000-year-old Neandertal individual using 454 sequencing. The high-throughput sequencing identified 8341 sequences of mtDNA from 4.8 GB of genetic material obtained from 0.2g of bone. The assembled sequence conclusively indicates that Neandertal mtDNA is distinct from modern human mtDNA and that the Neandertal mtDNA lineage diverged from the current human mtDNA lineage around 660,000 years ago. This means that the most recent common ancestor of Neandertal and human mtDNA lived

more than twice as long ago as the most recent common ancestor of current human mtDNAs (Green et al., 2008).

mtDNA analysis of a Hominin from South Siberia

Other than Neanderthals, whose DNA sequences have already been determined in a large number of individuals, the genetic linkages of other hominin lineages are mostly unknown. Despite recent developments in the extraction of aDNA from Neanderthals and early modern humans (Gilbert et al., 2007; Krause et al., 2010; Krings et al., 1997), DNA sequences have not been recovered from other Pleistocene hominins such as *H. erectus*, *H. heidelbergensis*, or *Homo antecessor*. One possible explanation for this is that in the case of ancient samples, the DNA was exposed to harsh environmental conditions for longer periods of time. DNA degradation accelerates with conditions like increased temperature and soil acidity (Smith et al., 2003).

A complete DNA sequence was retrieved from a bone unearthed in 2008 in Denisova Cave in southern Siberia using the Illumina platform. The mtDNA analysis showed that it represents an unknown kind of hominin mtDNA that shared a common ancestor with modern humans and Neanderthal mtDNAs around one million years ago. This suggests that it comes from a different hominin migration out of Africa than the forebears of Neanderthals and modern humans. The stratigraphic analysis of the cave where the bone was discovered implies that the Denisova hominin lived in close proximity to both Neanderthals and modern humans in both time and space (Krause et al., 2010).

mtDNA analysis of ancient mummies

The people of ancient Egypt believed in an afterlife. They preserved the integrity of their bodies through artificial mummification because in their belief the integrity of their physical bodies was important to continue their existence in the hereafter (Ikram, 2003). The mtDNA analysis using NGS technologies has been found to be instrumental in revealing the mysteries of these Egyptian mummies.

The first NGS-based mtDNA analysis of mummified samples that were obtained from 806 BCE to CE 124 was performed to trace the ancestral lineage of ancient Egyptians. The study characterized five Egyptian mummies and two excavated Bolivian lowland skeletons. In addition to human DNA, the sequencing libraries also included DNA from two parasites (Plasmodium and Toxoplasma), indicating the existence of the respective infections in mummies (Khairat et al., 2013).

Another study obtained 90 mitochondrial genomes and genome-wide datasets from three individuals from Egyptian mummies. The findings revealed that ancient Egyptians were more related to Near Easterners than modern-day Egyptians, who have received more Sub-Saharan mixing in recent times (Schuenemann et al., 2017).

In the last decade or so the entire mitochondrial genomes of the mummies have been sequenced, and their haplogroups have been identified (Gómez-Carballa et al., 2015). The oldest complete mitochondrial genome sequenced belonged to a prehistoric European, the Tyrolean Iceman, who lived around 5350–5100 years ago. The mtDNA obtained from the corpse was amplified and sequenced. The comparison of mtDNA with related extant lineages revealed that the Iceman belonged to the mitochondrial haplogroup K1 (Ermini et al., 2008).

Furthermore, the complete genome of a 7-year-old victim of an Inca child ritual that took place around 500 years ago. The mummy of the child was discovered in the Argentinian province of Mendoza. The analysis of mtDNA showed that the mummy belonged to the C1bi mitochondrial haplogroup, which matches the current haplogroups of Peru and Bolivia (Gómez-Carballa et al., 2015).

Overall, it appears that NGS technologies have sped up the mtDNA analysis of ancient mummies and paved the way for future research in this field. However, more extensive research is required in order to validate the findings obtained through archeology, literature, and anatomy.

Recent developments

The field of aDNA has been revolutionized by the rapid advancement in next generation technologies, which allow the amplification of very small aDNA fragments and yield a high amount of sequence data at a low cost. These NGS techniques have made it possible to detect misincorporations and have helped in the detection of modern exogenous DNA contaminations (Emery et al., 2022; Kirkinen et al., 2022). Such developments enhanced the authenticity of the aDNA analysis results.

The major challenge researchers faced in the analysis of NGS-produced data was its management. The data produced through high-throughput NGS technologies after aDNA analysis is in gigabytes and there was a need to harness advancements in bioinformatics and statistical tools to make the analysis of such huge data easier. Previously, bioinformatics tools such as mapDamage (Ginolhac et al., 2011) and mapDamage2.0 (Jónsson et al., 2013) were developed that helped in handling and identifying degraded sequences. Recently, a computational tool, damageProfiler, was developed that allows the detection of misincorporation in aDNA (Neukamm et al., 2021). Tools such as PMD tools were also used to describe damage patterns by identifying damage sequences that made decontamination of the aDNA possible (Skoglund et al., 2014). Current advancement in algorithmic tools, however, have allowed researchers to estimate exogenous contamination in human mtDNA analysis in ancient samples. These tools, ContamLD (Nakatsuka et al., 2020) and AuthentiCT (Peyrégne & Peter, 2020) coupled with previously designed tool, schmutzi (Renaud et al., 2015) have enabled scientists to reconstruct the endogenous consensus sequences with more accurately. Haplocheck

is another computational tool that employs mitochondrial phylogeny to differentiate between contaminated and endogenous DNA (Weissensteiner et al., 2021).

Some of the other important methodological contributions in aDNA analysis include the development of precise algorithms for the taxonomic assignment of aDNA data (Cribdon et al., 2020). The reconstruction of methylation signatures from DNA damage in ancient plant remains (Wagner et al., 2020) and the characterization of ancient genomes and methylomes using improved reference alignment of DNA obtained from ancient samples (Poullet & Orlando, 2020). Moreover, a laboratory information management system known as CASCADE has been developed for archiving and storing experimental data, including information related to preparation, extraction protocols, library construction, amplification, and sequencing of aDNA (Dolle et al., 2020). Recently, scientists have used a multiple-taxon approach to successfully integrate genomic data from multiple taxa to provide a more clear the understanding of interaction of early humans with animals and the environment (Dussex et al., 2021).

The present techniques available for aDNA research have significantly increased the accessibility of aDNA and opened doors for novel research. However, more research is needed to develop and optimize protocols for every step of aDNA analysis, especially in the area of hybridization enrichment. Some of the applications require available sequence data to become more precise and accurate, for example, mapping of NGS reads to a reference sequence; therefore, more and more sequence information needs to be collected from ancient samples. The computational tools also need to be improved to analyze and compare the huge amount of sequence data retrieved from ancient samples. Given these technological advancements and developments the direction of aDNA research is going to change radically in the coming years.

Future perspective

Maternal genome analysis has been proven an excellent source in forensic investigational and evolutionary studies where the sample is highly degraded and other genomic markers, such as autosomal DNA markers or paternal DNA markers fail to generate readable data. Analysis of the aDNA has been a source to get insight into the events of the past, such as migratory events, historically significant events, and human evolution. Advances in sequencing technology and the use of high-throughput techniques such as NGS have revolutionized the development. Now a bulk of data can be obtained and analyzed with much ease. Despite such developments, there is still room for development. Methods of ancient sample processing must be further improved to lower the permanent damage to the unique sample. Further, genomic DNA extraction must be optimized to enhance the yield of mtDNA. The amount of reagents used must be studied, and their role in noise production and experiment failure in NGS data analysis must be determined. Probe design for mtDNA capturing can be optimized for the

mtDNA control region, and mtDNA STRs can also be targeted for capturing mtDNA probes.

Every day, research has increased the understanding of haplotypes and haplogroupings of mtDNA. More knowledge in the field has great potential for getting a comprehensive understanding of human history. The formation of a database for aDNA data obtained through maternal genome analysis using NGS may have applications in unrevealing interand intrapopulation similarities and differences and linking that knowledge to historical migrations. Such information can also facilitate delineating population-specific mtDNA signatures that might have a potential role in forensic casework or in the design of population-specific medication for diseases.

References

Barnes, I., Matheus, P., Shapiro, B., Jensen, D., & Cooper, A. (2002). Dynamics of pleistocene population extinctions in Beringian brown bears. *Science, 295*(5563), 2267–2270. https://doi.org/10.1126/science.1067814

Basler, C. F., Reid, A. H., Dybing, J. K., Janczewski, T. A., Fanning, T. G., Zheng, H., Salvatore, M., Perdue, M. L., Swayne, D. E., García-Sastre, A., Palese, P., & Taubenberger, J. K. (2001). Sequence of the 1918 pandemic influenza virus nonstructural gene (NS) segment and characterization of recombinant viruses bearing the 1918 NS genes. *Proceedings of the National Academy of Sciences of the United States of America, 98*(5), 2746–2751. https://doi.org/10.1073/pnas.031575198

Behar, D. M., Van Oven, M., Rosset, S., Metspalu, M., Loogväli, E. L., Silva, N. M., Kivisild, T., Torroni, A., & Villems, R. (2012). A \copernican\ reassessment of the human mitochondrial DNA tree from its root. *The American Journal of Human Genetics, 90*(4), 675–684. https://doi.org/10.1016/j.ajhg.2012.03.002

Briggs, A. W., & Heyn, P. (2012). Preparation of next-generation sequencing libraries from damaged DNA. *Methods in Molecular Biology, 840*, 143–154. https://doi.org/10.1007/978-1-61779-516-9_18

Brück, J. (2021). Ancient DNA, kinship and relational identities in bronze age Britain. *Antiquity, 95*(379), 228–237. https://doi.org/10.15184/aqy.2020.216

Brüniche–Olsen, A., Jones, M. E., Burridge, C. P., Murchison, E. P., Holland, B. R., & Austin, J. J. (2018). Ancient DNA tracks the mainland extinction and island survival of the Tasmanian devil. *Journal of Biogeography, 45*(5), 963–976. https://doi.org/10.1111/jbi.13214

Calloway, C. D., Kim, H., & Erlich, H. (2015). *Resolution of DNA mixtures and analysis of degraded DNA using the 454 DNA sequencing technology*. DOJ Report.

Caramelli, D., Milani, L., Vai, S., Modi, A., Pecchioli, E., Girardi, M., Pilli, E., Lari, M., Lippi, B., Ronchitelli, A., Mallegni, F., Casoli, A., Bertorelle, G., Barbujani, G., & Harpending, H. (2008). A 28,000 Years old cro-magnon mtDNA sequence differs from all potentially contaminating modern sequences. *PLoS One, 3*(7), e2700. https://doi.org/10.1371/journal.pone.0002700

Carlyle, S. W., Parr, R. L., Hayes, M. G., & O'Rourke, D. H. (2000). Context of maternal lineages in the greater Southwest. *American Journal of Physical Anthropology, 113*(1), 85–101. https://doi.org/10.1002/1096-8644(200009)113:1<85::aid-ajpa8>3.0.co;2-1

Chyleński, M., Ehler, E., Somel, M., Yaka, R., Krzewińska, M., Dabert, M., Juras, A., & Marciniak, A. (2019). Ancient mitochondrial genomes reveal the absence of maternal kinship in the burials of Çatalhöyük people and their genetic affinities. *Genes, 10*(3), 207. https://doi.org/10.3390/genes10030207

Cole, L. W. (2016). The evolution of per-cell organelle number. *Frontiers in Cell and Developmental Biology, 4*. https://doi.org/10.3389/fcell.2016.00085

Cooper, A., Lalueza-Fox, C., Anderson, S., Rambaut, A., Austin, J., & Ward, R. (2001). Complete mitochondrial genome sequences of two extinct Moas clarify ratite evolution. *Nature, 409*(6821), 704–707. https://doi.org/10.1038/35055536

Cribdon, B., Ware, R., Smith, O., Gaffney, V., & Allaby, R. G. (2020). Pia: More accurate taxonomic assignment of metagenomic data demonstrated on sedaDNA from the north sea. *Frontiers in Ecology and Evolution, 8*. https://doi.org/10.3389/fevo.2020.00084

Cui, Y., Lindo, J., Hughes, C. E., Johnson, J. W., Hernandez, A. G., Kemp, B. M., Ma, J., Cunningham, R., Petzelt, B., Mitchell, J., Archer, D., Cybulski, J. S., Malhi, R. S., & Gilbert, M. T. P. (2013). Ancient DNA analysis of mid-holocene individuals from the Northwest coast of north America reveals different evolutionary paths for mitogenomes. *PLoS One, 8*(7), e66948. https://doi.org/10.1371/journal.pone.0066948

Dabney, J., Knapp, M., Glocke, I., Gansauge, M.-T., Weihmann, A., Nickel, B., Valdiosera, C., Garcia, N., Paabo, S., Arsuaga, J.-L., & Meyer, M. (2013). Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proceedings of the National Academy of Sciences, 110*(39), 15758−15763. https://doi.org/10.1073/pnas.1314445110

Damgaard, P. B., Margaryan, A., Schroeder, H., Orlando, L., Willerslev, E., & Allentoft, M. E. (2015). Improving access to endogenous DNA in ancient bones and teeth. *Scientific Reports, 5*. https://doi.org/10.1038/srep11184

Dolle, D., Fages, A., Mata, X., Schiavinato, S., Tonasso-Calvière, L., Chauvey, L., Wagner, S., Der Sarkissian, C., Fromentier, A., Seguin-Orlando, A., & Orlando, L. (2020). Cascade: A custom-made archiving system for the conservation of ancient DNA experimental data. *Frontiers in Ecology and Evolution, 8*. https://doi.org/10.3389/fevo.2020.00185

Donoghue, H. D., Spigelman, M., Greenblatt, C. L., Lev-Maor, G., Kahila Bar-Gal, G., Matheson, C., Vernon, K., Nerlich, A. G., & Zink, A. R. (2004). *Tuberculosis*: from prehistory to Robert Koch, as revealed by ancient DNA. *The Lancet Infectious Diseases, 4*(9), 584−592. https://doi.org/10.1016/S1473-3099(04)01133-8

Donoghue, H. D., Spigelman, M., Zias, J., Gernaey-Child, A. M., & Minnikin, D. E. (1998). *Mycobacterium tuberculosis* complex DNA in calcified pleura from remains 1400 years old. *Letters in Applied Microbiology, 27*(5), 265−269. https://doi.org/10.1046/j.1472-765X.1998.00436.x

Drancourt, M., & Raoult, D. (2002). Molecular insights into the history of plague. *Microbes and Infection, 4*(1), 105−109. https://doi.org/10.1016/S1286-4579(01)01515-5

Drosou, K., Price, C., & Brown, T. A. (2018). The kinship of two 12th Dynasty mummies revealed by ancient DNA sequencing. *Journal of Archaeological Science: Report, 17*, 793−797. https://doi.org/10.1016/j.jasrep.2017.12.025

Dussex, N., Bergfeldt, N., de Anca Prado, V., Dehasque, M., Díez-del-Molino, D., Ersmark, E., Kanellidou, F., Larsson, P., Lemež, Š., Lord, E., Mármol-Sánchez, E., Meleg, I. N., Måsviken, J., Naidoo, T., Studerus, J., Vicente, M., von Seth, J., Götherström, A., Dalén, L., & Heintzman, P. D. (2021). Integrating multi-taxon palaeogenomes and sedimentary ancient DNA to study past ecosystem dynamics. *Proceedings of the Royal Society B: Biological Sciences, 288*(1957). https://doi.org/10.1098/rspb.2021.1252

Emery, M. V., Bolhofner, K., Ghafoor, S., Winingear, S., Buikstra, J. E., Fulginiti, L. C., & Stone, A. C. (2022). Whole mitochondrial genomes assembled from thermally altered forensic bones and teeth. *Forensic Science International: Genetics, 56*, 102610. https://doi.org/10.1016/j.fsigen.2021.102610

Ermini, L., Olivieri, C., Rizzi, E., Corti, G., Bonnal, R., Soares, P., Luciani, S., Marota, I., De Bellis, G., Richards, M. B., & Rollo, F. (2008). Complete mitochondrial genome sequence of the Tyrolean iceman. *Current Biology, 18*(21), 1687−1693. https://doi.org/10.1016/j.cub.2008.09.028

Eskimo, M. D. J. A. (1984). *Handbook of north American Indians* (Vol. 5, pp. 347−358).

Etherington, G. J., Ramirez-Gonzalez, R. H., & MacLean, D. (2015). Bio-samtools 2: A package for analysis and visualization of sequence and alignment data with SAMtools in ruby. *Bioinformatics, 31*(15), 2565−2567. https://doi.org/10.1093/bioinformatics/btv178

Farrer, A. G., Wright, S. L., Skelly, E., Eisenhofer, R., Dobney, K., & Weyrich, L. S. (2021). Effectiveness of decontamination protocols when analyzing ancient DNA preserved in dental calculus. *Scientific Reports, 11*(1). https://doi.org/10.1038/s41598-021-86100-w

Fulton, T. L. (2012). Setting up an ancient DNA laboratory. *Methods in Molecular Biology, 840*, 1−11. https://doi.org/10.1007/978-1-61779-516-9_1

Furtwängler, A., Reiter, E., Neumann, G. U., Siebke, I., Steuri, N., Hafner, A., Lösch, S., Anthes, N., Schuenemann, V. J., & Krause, J. (2018). Ratio of mitochondrial to nuclear DNA affects contamination estimates in ancient DNA analysis. *Scientific Reports, 8*(1). https://doi.org/10.1038/s41598-018-32083-0

Gad, Y. Z., Hassan, N. A.-M., Mousa, D. M., Fouad, F. A., El-Sayed, S. G., Abdelazeem, M. A., Mahdy, S. M., Othman, H. Y., Ibrahim, D. W., Khairat, R., & Ismail, S. (2021). Insights from ancient DNA analysis of Egyptian human mummies: Clues to disease and kinship. *Human Molecular Genetics, 30*(R1), R24−R28. https://doi.org/10.1093/hmg/ddaa223

Gamba, C., Jones, E. R., Teasdale, M. D., McLaughlin, R. L., Gonzalez-Fortes, G., Mattiangeli, V., Domboróczki, L., Kővári, I., Pap, I., Anders, A., Whittle, A., Dani, J., Raczky, P., Higham, T. F. G., Hofreiter, M., Bradley, D. G., & Pinhasi, R. (2014). Genome flux and stasis in a five millennium transect of European prehistory. *Nature Communications, 5*(1). https://doi.org/10.1038/ncomms6257

Garrison, E., & Marth, G. (2012). *Haplotype-based variant detection from short-read sequencing.*

Geigl, E.-M., & Grange, T. (2018). Ancient DNA: The quest for the best. *Molecular Ecology Resources, 18*(6), 1185−1187. https://doi.org/10.1111/1755-0998.12931

Gibbons, A. (2021). When modern humans met Neanderthals. *Science, 372*(6538), 115−116. https://doi.org/10.1126/science.372.6538.115

Gilbert, M. T. P., Drautz, D. I., Lesk, A. M., Ho, S. Y. W., Qi, J., Ratan, A., Hsu, C. H., Sher, A., Dalén, L., Götherström, A., Tomsho, L. P., Rendulic, S., Packard, M., Campos, P. F., Kuznetsova, T. V., Shidlovskiy, F., Tikhonov, A., Willerslev, E., Iacumin, P., … Schuster, S. C. (2008). Intraspecific phylogenetic analysis of Siberian woolly mammoths using complete mitochondrial genomes. *Proceedings of the National Academy of Sciences of the United States of America, 105*(24), 8327−8332. https://doi.org/10.1073/pnas.0802315105

Gilbert, M. T. P., Kivisild, T., Grønnow, B., Andersen, P. K., Metspalu, E., Reidla, M., Tamm, E., Axelsson, E., Götherström, A., Campos, P. F., Rasmussen, M., Metspalu, M., Higham, T. F. G., Schwenninger, J. L., Nathan, R., De Hoog, C. J., Koch, A., Møller, L. N., Andreasen, C., … Willerslev, E. (2008). Paleo-Eskimo mtDNA genome reveals matrilineal discontinuity in Greenland. *Science, 320*(5884), 1787−1789. https://doi.org/10.1126/science.1159750

Gilbert, M. T. P., Rudbeck, L., Willerslev, E., Hansen, A. J., Smith, C., Penkman, K. E. H., Prangenberg, K., Nielsen-Marsh, C. M., Jans, M. E., Arthur, P., Lynneruo, N., Turner-Walker, G., Biddle, M., Kjølbye, B., & Collins, M. J. (2005). Biochemical and physical correlates of DNA contamination in archaeological human bones and teeth excavated at Matera, Italy. *Journal of Archaeological Science, 32*(5), 785−793. https://doi.org/10.1016/j.jas.2004.12.008

Gilbert, M. T. P., Tomsho, L. P., Rendulic, S., Packard, M., Drautz, D. I., Sher, A., Tikhonov, A., Dalén, L., Kuznetsova, T., Kosintsev, P., Campos, P. F., Higham, T., Collins, M. J., Wilson, A. S., Shidlovskiy, F., Buigues, B., Ericson, P. G. P., Germonpré, M., Götherström, A., … Schuster, S. C. (2007). Whole-genome shotgun sequencing of mitochondria from ancient hair shafts. *Science, 317*(5846), 1927−1930. https://doi.org/10.1126/science.1146971

Gilbert, M. T. P., Willerslev, E., Hansen, A. J., Barnes, I., Rudbeck, L., Lynnerup, N., & Cooper, A. (2003). Distribution patterns of postmortem damage in human mitochondrial DNA. *The American Journal of Human Genetics, 72*(1), 32−47. https://doi.org/10.1086/345378

Gilbert, M. T. P., Wilson, A. S., Bunce, M., Hansen, A. J., Willerslev, E., Shapiro, B., Higham, T. F. G., Richards, M. P., O'Connell, T. C., Tobin, D. J., Janaway, R. C., & Cooper, A. (2004). Ancient mitochondrial DNA from hair [1]. *Current Biology, 14*(12), R463−R464. https://doi.org/10.1016/j.cub.2004.06.008

Ginolhac, A., Rasmussen, M., Gilbert, M. T. P., Willerslev, E., & Orlando, L. (2011). mapDamage: Testing for damage patterns in ancient DNA sequences. *Bioinformatics, 27*(15), 2153−2155. https://doi.org/10.1093/bioinformatics/btr347

Gómez-Carballa, A., Catelli, L., Pardo-Seco, J., Martinón-Torres, F., Roewer, L., Vullo, C., & Salas, A. (2015). The complete mitogenome of a 500-year-old Inca child mummy. *Scientific Reports, 5*(1). https://doi.org/10.1038/srep16462

Green, R. E., Briggs, A. W., Krause, J., Prüfer, K., Burbano, H. A., Siebauer, M., Lachmann, M., & Pääbo, S. (2009). The Neandertal genome and ancient DNA authenticity. *EMBO Journal, 28*(17), 2494–2502. https://doi.org/10.1038/emboj.2009.222

Green, R. E., Krause, J., Ptak, S. E., Briggs, A. W., Ronan, M. T., Simons, J. F., Du, L., Egholm, M., Rothberg, J. M., Paunovic, M., & Pääbo, S. (2006). Analysis of one million base pairs of Neanderthal DNA. *Nature, 444*(7117), 330–336. https://doi.org/10.1038/nature05336

Green, R. E., Malaspinas, A. S., Krause, J., Briggs, A. W., Johnson, P. L. F., Uhler, C., Meyer, M., Good, J. M., Maricic, T., Stenzel, U., Prüfer, K., Siebauer, M., Burbano, H. A., Ronan, M., Rothberg, J. M., Egholm, M., Rudan, P., Brajković, D., Kućan, Z., … Pääbo, S. (2008). A complete Neandertal mitochondrial genome sequence determined by high-throughput sequencing. *Cell, 134*(3), 416–426. https://doi.org/10.1016/j.cell.2008.06.021

Grigorenko, A. P., Borinskaya, S. A., Yankovsky, N. K., & Rogaev, E. I. (2009). Achievements and peculiarities in studies of ancient DNA and DNA from complicated forensic specimens. *Acta Nature, 1*(3), 58–69. https://doi.org/10.32607/20758251-2009-1-3-58-69

Harney, É., Cheronet, O., Fernandes, D. M., Sirak, K., Mah, M., Bernardos, R., Adamski, N., Broomandkhoshbacht, N., Callan, K., Lawson, A. M., Oppenheimer, J., Stewardson, K., Zalzala, F., Anders, A., Candilio, F., Constantinescu, M., Coppa, A., Ciobanu, I., Dani, J., … Pinhasi, R. (2021). A minimally destructive protocol for DNA extraction from ancient teeth. *Genome Research, 31*(3), 472–483. https://doi.org/10.1101/GR.267534.120

Hawks, J. (2021). Accurate depiction of uncertainty in ancient DNA research: The case of Neandertal ancestry in Africa. *Journal of Social Archaeology, 21*(2), 179–196. https://doi.org/10.1177/1469605321995616

Hofreiter, M., Serre, D., Poinar, H.N., Kuch, M., & Pääbo, S. (2001). Ancient DNA. *Nature Reviews Genetics*, 2(5), 353–359.

Hofreiter, M., Jaenicke, V., Serre, D., Von Haeseler, A., & Pääbo, S. (2001). DNA sequences from multiple amplifications reveal artifacts induced by cytosine deamination in ancient DNA. *Nucleic Acids Research, 29*(23), 4793–4799. https://doi.org/10.1093/nar/29.23.4793

Hofreiter, M., Poinar, H. N., Spaulding, W. G., Bauer, K., Martin, P. S., Possnert, G., & Pääbo, S. (2000). A molecular analysis of ground sloth diet through the last glaciation. *Molecular Ecology, 9*(12), 1975–1984. https://doi.org/10.1046/j.1365-294X.2000.01106.x

Hofreiter, M., Sneberger, J., Pospisek, M., & Vanek, D. (2021). Progress in forensic bone DNA analysis: Lessons learned from ancient DNA. *Forensic Science International: Genetics, 54*, 102538. https://doi.org/10.1016/j.fsigen.2021.102538

Ikram, S. (2003). *Death and burial in ancient Egypt*. Longman Publishing Group.

Jaenicke-Després, V., Buckler, E. S., Smith, B. D., Gilbert, M. T. P., Cooper, A., Doebley, J., & Pääbo, S. (2003). Early allelic selection in maize as revealed by ancient DNA. *Science, 302*(5648), 1206–1208. https://doi.org/10.1126/science.1089056

Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P. L. F., & Orlando, L. (2013). MapDamage2.0: Fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics, 29*(13), 1682–1684. https://doi.org/10.1093/bioinformatics/btt193

Jordan, D., & Mills, D. (2021). Past, present, and future of DNA typing for analyzing human and non-human forensic samples. *Frontiers in Ecology and Evolution, 9*. https://doi.org/10.3389/fevo.2021.646130

Keyser-Tracqui, C., Crubézy, E., & Ludes, B. (2003). Nuclear and mitochondrial DNA analysis of a 2,000-year-old necropolis in the Egyin Gol valley of Mongolia. *The American Journal of Human Genetics, 73*(2), 247–260. https://doi.org/10.1086/377005

Khairat, R., Ball, M., Chang, C. C. H., Bianucci, R., Nerlich, A. G., Trautmann, M., Ismail, S., Shanab, G. M. L., Karim, A. M., Gad, Y. Z., & Pusch, C. M. (2013). First insights into the metagenome of Egyptian mummies using next-generation sequencing. *Journal of Applied Genetics, 54*(3), 309–325. https://doi.org/10.1007/s13353-013-0145-1

Khan, K., Siddiqi, M. H., Abbas, M., Almas, M., & Idrees, M. (2017). Forensic applications of Y chromosomal properties. *Legal Medicine, 26*, 86–91. https://doi.org/10.1016/j.legalmed.2017.04.002

Khan, K., Siddiqi, M. H., Ali, S., Naqvi, A.-U.-N., Ali, S., & Sabar, M. F. (2020). Mitochondrial DNA control region variants analysis in Balti population of Gilgit-Baltistan, Pakistan. *Meta Gene, 23*, 100630. https://doi.org/10.1016/j.mgene.2019.100630

Kirkinen, T., Honka, J., Salazar, D., Kvist, L., Saastamoinen, M., & Hemmann, K. (2022). Determination of different predictors affecting DNA concentration isolated from historical hairs of the Finnhorse. *Journal of Archaeological Science: Report, 41*, 103262. https://doi.org/10.1016/j.jasrep.2021.103262

Knapp, M., Clarke, A. C., Horsburgh, K. A., & Matisoo-Smith, E. A. (2012). Setting the stage − building and working in an ancient DNA laboratory. *Annals of Anatomy - Anatomischer Anzeiger, 194*(1), 3−6. https://doi.org/10.1016/j.aanat.2011.03.008

Kolman, C. J. (1999). *Molecular anthropology* (pp. 183−200). Springer Science and Business Media LLC. https://doi.org/10.1007/978-1-4615-4263-6_13

Krause, J., Briggs, A. W., Kircher, M., Maricic, T., Zwyns, N., Derevianko, A., & Pääbo, S. (2010). A complete mtDNA genome of an early modern human from Kostenki, Russia. *Current Biology, 20*(3), 231−236. https://doi.org/10.1016/j.cub.2009.11.068

Krause, J., Fu, Q., Good, J. M., Viola, B., Shunkov, M. V., Derevianko, A. P., & Pääbo, S. (2010). The complete mitochondrial DNA genome of an unknown hominin from southern Siberia. *Nature, 464*(7290), 894−897. https://doi.org/10.1038/nature08976

Krings, M., Capelli, C., Tschentscher, F., Geisert, H., Meyer, S., Von Haeseler, A., Grossschmidt, K., Possnert, G., Paunovic, M., & Pääbo, S. (2000). A view of Neandertal genetic diversity. *Nature Genetics, 26*(2), 144−146. https://doi.org/10.1038/79855

Krings, M., Stone, A., Schmitz, R. W., Krainitzki, H., Stoneking, M., & Pääbo, S. (1997). Neandertal DNA sequences and the origin of modern humans. *Cell, 90*(1), 19−30. https://doi.org/10.1016/S0092-8674(00)80310-4

Lalueza, C., Pérez-Pérez, A., Prats, E., Cornudella, L., & Turbón, D. (1997). Lack of founding Amerindian mitochondrial DNA lineages in extinct Aborigines from Tierra del Fuego-Patagonia. *Human Molecular Genetics, 6*(1), 41−46. https://doi.org/10.1093/hmg/6.1.41

Leonard, J. A., Wayne, R. K., Wheeler, J., Valadez, R., Guillén, S., & Vilà, C. (2002). Ancient DNA evidence for old world origin of new world dogs. *Science, 298*(5598), 1613−1616. https://doi.org/10.1126/science.1076980

Lindahl, T. (1993). Instability and decay of the primary structure of DNA. *Nature, 362*(6422), 709−715. https://doi.org/10.1038/362709a0

Lindahl, T., & Karlstrom, O. (1973). Heat-induced depyrimidination of deoxyribonucleic acid in neutral solution. *Biochemistry, 12*(25), 5151−5154. https://doi.org/10.1021/bi00749a020

Lindahl, T., & Nyberg, B. (1972). Rate of depurination of native deoxyribonucleic acid. *Biochemistry, 11*(19), 3610−3618. https://doi.org/10.1021/bi00769a018

Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A., Berka, J., Braverman, M. S., Chen, Y. J., Chen, Z., Dewell, S. B., Du, L., Fierro, J. M., Gomes, X. V., Godwin, B. C., He, W., Helgesen, S., Ho, C. H., Irzyk, G. P., … Rothberg, J. M. (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature, 437*(7057), 376−380. https://doi.org/10.1038/nature03959

Matisoo-Smith, E., Allen, J. S., Ladefoged, T. N., Roberts, R. M., & Lambert, D. M. (1997). Ancient DNA from polynesian rats: Extraction, amplification and sequence from single small bones. *Electrophoresis, 18*(9), 1534−1537. https://doi.org/10.1002/elps.1150180908. Wiley-VCH Verlag.

Matsvay, A. D., Alborova, I. E., Pimkina, E. V., Markelov, M. L., Khafizov, K., & Mustafin, K. K. (2019). Experimental approaches for ancient DNA extraction and sample preparation for next generation sequencing in ultra-clean conditions. *Conservation Genetics Resources, 11*(3), 345−353. https://doi.org/10.1007/s12686-018-1016-1

Millar, C. D., Huynen, L., Subramanian, S., Mohandesan, E., & Lambert, D. M. (2008). New developments in ancient genomics. *Trends in Ecology & Evolution, 23*(7), 386−393. https://doi.org/10.1016/j.tree.2008.04.002

Miller, W., Drautz, D. I., Janecka, J. E., Lesk, A. M., Ratan, A., Tomsho, L. P., Packard, M., Zhang, Y., McClellan, L. R., Qi, J., Zhao, F., Gilbert, M. T. P., Dalén, L., Arsuaga, J. L., Ericson, P. G. P., Huson, D. H., Helgen, K. M., Murphy, W. J., Götherström, A., & Schuster, S. C. (2009). The

mitochondrial genome sequence of the Tasmanian tiger [*Thylacinus cynocephalus*]. *Genome Research, 19*(2), 213–220. https://doi.org/10.1101/gr.082628.108

Modi, A., Tassi, F., Susca, R. R., Vai, S., Rizzi, E., Bellis, G. D., Lugliè, C., Gonzalez Fortes, G., Lari, M., Barbujani, G., Caramelli, D., & Ghirotto, S. (2017). Complete mitochondrial sequences from mesolithic sardinia. *Scientific Reports, 7*(1). https://doi.org/10.1038/srep42869

Mohni, K. N., Wessel, S. R., Zhao, R., Wojciechowski, A. C., Luzwick, J. W., Layden, H., Eichman, B. F., Thompson, P. S., Mehta, K. P. M., & Cortez, D. (2019). HMCES maintains genome integrity by shielding abasic sites in single-strand DNA. *Cell, 176*(1–2), 144–153.e13. https://doi.org/10.1016/j.cell.2018.10.055

Mudge, C., Gray, L. J., & Austin, J. J. (2021). Using mitochondrial DNA to identify the provenance of 19th century Kākāpō skins held in Australia's oldest natural history collection, the Macleay. *Emu - Austral Ornithology, 1–11*. https://doi.org/10.1080/01584197.2021.1998782

Nakatsuka, N., Harney, É., Mallick, S., Mah, M., Patterson, N., & Reich, D. (2020). ContamLD: Estimation of ancient nuclear DNA contamination using breakdown of linkage disequilibrium. *Genome Biology, 21*(1). https://doi.org/10.1186/s13059-020-02111-2

Neparáczki, E., Kocsy, K., Tóth, G. E., Maróti, Z., Kalmár, T., Bihari, P., Nagy, I., Pálfi, G., Molnár, E., Raskó, I., Török, T., & Caramelli, D. (2017). Revising mtDNA haplotypes of the ancient Hungarian conquerors with next generation sequencing. *PLoS One, 12*(4), e0174886. https://doi.org/10.1371/journal.pone.0174886

Neukamm, J., Peltzer, A., Nieselt, K., & Kelso, J. (2021). DamageProfiler: Fast damage pattern calculation for ancient DNA. *Bioinformatics, 37*(20), 3652–3653. https://doi.org/10.1093/bioinformatics/btab190

Nistelberger, H. M., Pálsdóttir, A. H., Star, B., Leifsson, R., Gondek, A. T., Orlando, L., Barrett, J. H., Hallsson, J. H., & Boessenkool, S. (2019). Sexing Viking Age horses from burial and non-burial sites in Iceland using ancient DNA. *Journal of Archaeological Science, 101*, 115–122. https://doi.org/10.1016/j.jas.2018.11.007

Noonan, J. P., Coop, G., Kudaravalli, S., Smith, D., Krause, J., Alessi, J., Chen, F., Platt, D., Pääbo, S., Pritchard, J. K., & Rubin, E. M. (2006). Sequencing and analysis of Neanderthal genomic DNA. *Science, 314*(5802), 1113–1118. https://doi.org/10.1126/science.1131412

Orlando, L., Bonjean, D., Bocherens, H., Thenot, A., Argant, A., Otte, M., & Hänni, C. (2002). Ancient DNA and the population genetics of cave bears (*Ursus spelaeus*) through space and time. *Molecular Biology and Evolution, 19*(11), 1920–1933. https://doi.org/10.1093/oxfordjournals.molbev.a004016

Parson, W., Strobl, C., Huber, G., Zimmermann, B., Gomes, S. M., Souto, L., Fendt, L., Delport, R., Langit, R., Wootton, S., Lagacé, R., & Irwin, J. (2013). Evaluation of next generation mtGenome sequencing using the Ion Torrent personal genome machine (PGM). *Forensic Science International: Genetics, 7*(5), 543–549. https://doi.org/10.1016/j.fsigen.2013.06.003

Peyrégne, S., & Peter, B. M. (2020). AuthentiCT: A model of ancient DNA damage to estimate the proportion of present-day DNA contamination. *Genome Biology, 21*(1). https://doi.org/10.1186/s13059-020-02123-y

Pinhasi, R., Fernandes, D. M., Sirak, K., & Cheronet, O. (2019). Isolating the human cochlea to generate bone powder for ancient DNA analysis. *Nature Protocols, 14*(4), 1194–1205. https://doi.org/10.1038/s41596-019-0137-7

Poinar, H. N., & Cooper, A. J. S. (2000). Ancient DNA: Do it right or not at all. *Science, 5482*.

Poinar, H. N., Hofreiter, M., Spaulding, W. G., Martin, P. S., Stankiewicz, B. A., Bland, H., Evershed, R. P., Possnert, G., & Pääbo, S. (1998). Molecular coproscopy: Dung and diet of the extinct ground sloth *Nothrotheriops shastensis*. *Science, 281*(5375), 402–406. https://doi.org/10.1126/science.281.5375.402

Poinar, H. N., Schwarz, C., Qi, J., Shapiro, B., MacPhee, R. D. E., Buigues, B., Tikhonov, A., Huson, D. M., Tomsho, L. P., Auch, A., Rampp, M., Miller, W., & Schuster, S. C. (2006). Metagenomics to paleogenomics: Large-scale sequencing of mammoth DNA. *Science, 311*(5759), 392–394. https://doi.org/10.1126/science.1123360

Poullet, M., & Orlando, L. (2020). Assessing DNA sequence alignment methods for characterizing ancient genomes and methylomes. *Frontiers in Ecology and Evolution, 8*. https://doi.org/10.3389/fevo.2020.00105

Renaud, G., Schubert, M., Sawyer, S., & Orlando, L. (2019). Authentication and assessment of contamination in ancient DNA. *Methods Molecular Biology, 1963*, 163–194. https://doi.org/10.1007/978-1-4939-9176-1_17. Springer Science and Business Media LLC.

Renaud, G., Slon, V., Duggan, A. T., & Kelso, J. (2015). Schmutzi: Estimation of contamination and endogenous mitochondrial consensus calling for ancient DNA. *Genome Biology, 16*(1). https://doi.org/10.1186/s13059-015-0776-0

Rizzi, E., Lari, M., Gigli, E., De Bellis, G., & Caramelli, D. (2012). Ancient DNA studies: New perspectives on old samples. *Genetics Selection Evolution, 44*(1). https://doi.org/10.1186/1297-9686-44-21

Rohland, N., Glocke, I., Aximu-Petri, A., & Meyer, M. (2018). Extraction of highly degraded DNA from ancient bones, teeth and sediments for high-throughput sequencing. *Nature Protocols, 13*(11), 2447–2461. https://doi.org/10.1038/s41596-018-0050-5

Rohland, N., & Hofreiter, M. (2007). Ancient dna extraction from bones and teeth. *Nature Protocols, 2*(7), 1756–1762. https://doi.org/10.1038/nprot.2007.247

Rohland, N., Siedel, H., & Hofreiter, M. (2010). A rapid column-based ancient DNA extraction method for increased sample throughput. *Molecular Ecology Resources, 10*(4), 677–683. https://doi.org/10.1111/j.1755-0998.2009.02824.x

Sampietro, M. L., Caramelli, D., Lao, O., Calafell, F., Comas, D., Lari, M., Agustí, B., Bertranpetit, J., & Lalueza-Fox, C. (2005). The genetics of the pre-roman Iberian Peninsula: A mtDNA study of ancient Iberians. *Annals of Human Genetics, 69*(5), 535–548. https://doi.org/10.1111/j.1529-8817.2005.00194.x

Sarmento, C. . A study on burrows-Wheeler aligner's performance optimization for ancient DNA mapping. In International **c**onference on practical applications of computational biology and bioinformatics. Vol 2021. Springer.

Schuenemann, V. J., Peltzer, A., Welte, B., Van Pelt, W. P., Molak, M., Wang, C. C., Furtwängler, A., Urban, C., Reiter, E., Nieselt, K., Teßmann, B., Francken, M., Harvati, K., Haak, W., Schiffels, S., & Krause, J. (2017). Ancient Egyptian mummy genomes suggest an increase of Sub-Saharan African ancestry in post-Roman periods. *Nature Communications, 8*. https://doi.org/10.1038/ncomms15694

Senovska, A., Drozdova, E., Vaculik, O., Pardy, F., Brzobohata, K., Fialova, D., Smerda, J., & Kos, P. (2021). Cost-effective straightforward method for captured whole mitogenome sequencing of ancient DNA. *Forensic Science International, 319*, 110638. https://doi.org/10.1016/j.forsciint.2020.110638

Shapiro, R. (1981). *Damage to DNA caused by hydrolysis* (pp. 3–18). Springer Science and Business Media LLC. https://doi.org/10.1007/978-1-4684-7956-0_1

Shendure, J., Mitra, R. D., Varma, C., & Church, G. M. (2004). Advanced sequencing technologies: Methods and goals. *Nature Reviews Genetics, 5*(5), 335–344. https://doi.org/10.1038/nrg1325

Shih, S., Bose, N., Gonçalves, A., Erlich, H., & Calloway, C. (2018). Applications of probe capture enrichment next generation sequencing for whole mitochondrial genome and 426 nuclear SNPs for forensically challenging samples. *Genes, 9*(1), 49. https://doi.org/10.3390/genes9010049

Skoglund, P., Northoff, B. H., Shunkov, M. V., Derevianko, A. P., Pääbo, S., Krause, J., & Jakobsson, M. (2014). Separating endogenous ancient DNA from modern day contamination in a Siberian Neandertal. *Proceedings of the National Academy of Sciences of the United States of America, 111*(6), 2229–2234. https://doi.org/10.1073/pnas.1318934111

Smith, C. I., Chamberlain, A. T., Riley, M. S., Stringer, C., & Collins, M. J. (2003). The thermal history of human fossils and the likelihood of successful DNA amplification. *Journal of Human Evolution, 45*(3), 203–217. https://doi.org/10.1016/S0047-2484(03)00106-4

Templeton, J. E. L., Brotherton, P. M., Llamas, B., Soubrier, J., Haak, W., Cooper, A., & Austin, J. J. (2013). DNA capture and next-generation sequencing can recover whole mitochondrial genomes from highly degraded samples for human identification. *Investigative Genetics, 4*(1). https://doi.org/10.1186/2041-2223-4-26

Vanscott, E. J., Ekel, T. M., & Auerbach, R. (1963). Determinants of rate and Kinetics of cell division in scalp hair. *Journal of Investigative Dermatology, 41*(5), 269–273. https://doi.org/10.1038/jid.1963.52

Vernot, B., Zavala, E. I., Gómez-Olivencia, A., Jacobs, Z., Slon, V., Mafessoni, F., Romagné, F., Pearson, A., Petr, M., Sala, N., Pablos, A., Aranburu, A., de Castro, J. M. B., Carbonell, E., Li, B., Krajcarz, M. T., Krivoshapkin, A. I., Kolobova, K. A., Kozlikin, M. B., … Meyer, M. (2021).

Unearthing Neanderthal population history using nuclear and mitochondrial DNA from cave sediments. *Science, 372*(6542). https://doi.org/10.1126/science.abf1667

Wagner, S., Plomion, C., & Orlando, L. (2020). Uncovering signatures of DNA methylation in ancient plant remains from patterns of post-mortem DNA damage. *Frontiers in Ecology and Evolution, 8*. https://doi.org/10.3389/fevo.2020.00011

Wall, J. D., & Kim, S. K. (2007). Inconsistencies in neanderthal genomic DNA sequences. *PLoS Genetics, 3*(10), 1862–1866. https://doi.org/10.1371/journal.pgen.0030175

Wasef, S., Wright, J. L., Adams, S., Westaway, M. C., Flinders, C., Willerslev, E., & Lambert, D. (2020). Insights into aboriginal Australian mortuary practices: Perspectives from ancient DNA. *Frontiers in Ecology and Evolution, 8*. https://doi.org/10.3389/fevo.2020.00217

Weissensteiner, H., Forer, L., Fendt, L., Kheirkhah, A., Salas, A., Kronenberg, F., & Schoenherr, S. (2021). Contamination detection in sequencing studies using the mitochondrial phylogeny. *Genome Research, 31*(2), 309–316. https://doi.org/10.1101/GR.256545.119

Willerslev, E., & Cooper, A. (2005). Ancient DNA. *Proceedings of the Royal Society B: Biological Sciences, 272*(1558), 3–16. https://doi.org/10.1098/rspb.2004.2813

Woods, R., Barnes, I., Brace, S., Turvey, S. T., & Ávila-Arcos, M. C. (2021). Ancient DNA suggests single colonization and Within-archipelago diversification of Caribbean caviomorph rodents. *Molecular Biology and Evolution, 38*(1), 84–95. https://doi.org/10.1093/molbev/msaa189

Xavier, C., Eduardoff, M., Bertoglio, B., Amory, C., Berger, C., Casas-Vargas, A., Pallua, J., & Parson, W. (2021). Evaluation of DNA extraction methods developed for forensic and ancient DNA applications using bone samples of different age. *Genes, 12*(2), 146. https://doi.org/10.3390/genes12020146

Zhang, H., Burr, S. P., & Chinnery, P. F. (2018). The mitochondrial DNA genetic bottleneck: Inheritance and beyond. *Essays in Biochemistry, 62*(3), 225–234. https://doi.org/10.1042/EBC20170096

This page intentionally left blank

CHAPTER 12

Application of NGS technology for parentage testing and relatedness analysis

Andrei Semikhodskii
Medical Genomics Ltd, London, United Kingdom

Introduction

Parentage and kinship testing through history

Throughout history, the question of biological kinship has been of crucial significance for human societies. From the mythical origin of Greek goods, some of whom, like Heracles or Theseus, had a mortal mother and two putative fathers (one mortal and one divine), to sacred texts of many religions, examples of questionable parentage and kinship are in abundance.

Over the years, various methods have been developed to resolve disputed parentage. In probably the most famous maternity dispute, King Solomon employed a psychological test of maternal love to identify the true mother of a child. As the story goes (1 Kings 3: 16–28) two women came to the king and stood before him claiming to be mothers to a newborn child. The King ordered to cut the living child in two and give half to one and a half to the other woman. Solomon's thinking was that the true mother of the baby would rather sacrifice her child's custody than her child's life. This is exactly what happened, and the King awarded the custody of the child to the woman, who, upon hearing the King's verdict, cried: "Please, my lord, give her the living baby! Don't kill him!".

Before the advent of modern serological and DNA-based methods in the 20th century parentage disputes were settled based on morphological, anthropological, and medical evidence. The exclusion of paternity was based on evidence of presumable impotence or ejaculation dysfunction of a potential father. The length of the gestation period, abnormal morphology of reproductive organs, various anthropological tests, and even the presence of certain characteristics in the child thought to be inherited from the alleged father (AF) were used as irrefutable evidence to prove or disprove questionable paternity.

The discovery of spermatozoa by Johan Ham and Antonie van Leeuwenhoek of Delft in 1677 (Albrecht & Schultheiss, 2004) and the human egg by Karl Ernst von Baer in 1827, followed by the observation by Oscar and Richard Hertwig in 1875–78 of the fusion of the egg's and spermatozoon's nuclei explained the nature of fertilization. These advances

Next Generation Sequencing (NGS) Technology in DNA Analysis
ISBN 978-0-323-99144-5, https://doi.org/10.1016/B978-0-323-99144-5.00008-1
© 2024 Elsevier Inc.
All rights reserved.

in embryology together with the discovery in the 1860s of the laws of heredity by Gregor Mendel laid a firm scientific foundation for modern parentage and kinship analysis.

The twentieth century saw the introduction of scientific methods that created contemporary genetic analysis. The first major step in the development of parentage and kinship testing based on the principles of heredity was the discovery of the blood groups by Karl Landsteiner in 1901 (Landsteiner, 1901). This was followed by the fundamental work by Von Dungern and Hirschfeld (1910) on the mendelian inheritance of the ABO blood groups, leading to their application to paternity suits by Reuben Ottenberg (Ottenberg, 1921) and marking the beginning of the era of forensic genetics. The MN and the Rh systems were added in 1927 and 1940, respectively. These three systems combined could exclude a wrongly accused male 50% of the time (Shaw, 1983). Now, decisions as to the true parentage of a child could be based on the results of molecular marker analysis considered within the framework of the laws of genetics. By 1974, there were 57 genetic systems (25 immunological and 32 biochemical) that could be used for parentage analysis (Chakraborty et al., 1974). For the first time, it became possible to achieve a probability of exclusion (PE) of 99% (Shaw, 1983) and calculate probabilities of paternity in nonexclusion cases.

The application of DNA analysis to paternity and kinship cases could be traced to the discovery of restriction enzymes and restriction fragment length polymorphism in the late 1970s, but the real breakthrough was done through the invention of the process of DNA fingerprinting by Sir Alec Jeffreys at the beginning of the 1980s (Gill et al., 1985) and its first application to resolve an immigration paternity dispute (Jeffries et al., 1985). The "Eureka moment" for the application of DNA analysis for human identity testing was the recognition by Sir Alec that repeating sequences, which are abundant in the human genome, are highly variable and informative genetic markers that can provide accurate individualizing information. Among the various types of repeat sequences found in the human genome microsatellites, known as short tandem repeats (STR) were established in the early 1990s as the main marker type for parentage and kinship analysis.

Approximately at the same time, DNA typing was augmented by the recently developed in vitro DNA amplification technology called the polymerase chain reaction (PCR) (Mullis et al., 1986). Combining STR markers with PCR amplification allowed the analysis of extremely small amounts of DNA (often as little as 10–15 pg) as well as DNA of various degrees of degradation. No more did DNA genotyping need to rely on the availability of large amounts of good-quality DNA, usually obtained from fresh blood. Blood and saliva stains, various tissue samples, buccal swabs, and even hair roots became routine sources of DNA for testing.

Currently, all parentage and kinship tests are regularly resolved by assessing 11–25 STR loci. According to the latest (2013) available report by the Association for the Advancement of Blood and Biotherapies (former American Association of Blood Banks, AABB) more than 99.9% of all paternity and kinship tests in the USA are performed using STR markers.

Principles of parentage and kinship analysis

Parentage analysis is based on the simple presumption that in the absence of mutation, a child must receive one allele matching each parent at every locus tested (Tracey, 2001). This allele is called an "obligate allele". When for all the molecular markers tested the child (C) and the alleged parent (AP) share matching obligate alleles, the parentage cannot be excluded and the weight of evidence in favor of the presumed parentage is calculated, taking into the account allelic frequencies of the markers analyzed. This weight of evidence is known as the combined parentage index (CPI). When a genetic inconsistency (a mismatch between the genotypes of AP and C) is observed it may be an indication of parentage exclusion; however, the possibility of the mismatch being caused by mutation must always be considered (Gjertson et al., 2007). If the CPI value of $> 10{,}000$ is reached the genetic evidence is thought to be in favor of parentage, and the mismatched alleles (if any) are presumed to be mutations (Li et al., 2017). Otherwise, analysis of additional markers is required. Mismatches of one or two alleles are commonly observed in parentage testing. Taking mutation events into account the CPI when two mutations are suspected is very low, bordering on single digits. When three inconsistencies have been detected, the CPI is usually close to zero, and this is deemed sufficient to issue a paternity exclusion report (Gjertson et al., 2007). At the same time, nonexclusion cases with three mutations have been described (Jacewicz et al., 2004), and if such a situation is suspected, a major effort to find and test additional markers is required in order not to come to a false-negative conclusion.

Kinship testing, other than parentage, is based on a different presumption. Here, no obligate alleles are considered, but instead, a probability that alleles identical by state (IBS) between the individuals tested are also identical by descent (IBD) (i.e., originate from a common relative(s) of these individuals) as opposed to being identical by chance (i.e., inherited from unrelated ancestors) is evaluated. The statistical evidence in kinship testing is less strong than in parentage analysis. It is theoretically impossible to exclude kinship even when inconsistencies are observed at each locus tested. Whereas in most cases, parentage exclusion can be stated with more or less certainty, unlikely kinship is stated as having a low probability, and likely kinship is stated as having a high probability (Wenk, 2004). Even though modern STR multiplexes contain 25 loci or more, it is sometimes problematic to determine kinship between first-degree relatives (full siblings) and even more difficult between second-degree relatives (half-siblings, grandparent-grandchild, avuncular relationship, etc).

Samples used for parentage and kinship analysis

Until the mid-1990s, blood was the main sample for parentage and kinship investigations. These days, with the introduction of PCR and the development of capillary electrophoresis (CE), buccal (mouth) swabs have superseded blood as the sample of choice. According to the latest available report of the AABB Relationship Testing Standards Program

Unit (2013), buccal swabs constituted more than 99.5% of all samples used for relationship testing in the USA. Buccal swabs provide good-quality DNA and generally do not cause complications in the analysis and interpretation of the results. However, in certain cases, samples from decomposed bodies, histological samples, or pregnancy termination products have to be used for DNA analysis. These samples are not ideal and pose considerable technical challenges, which often affect the analysis and interpretation of the data. DNA from decomposed bodies and histological samples is commonly of inadequate quality. The fixation process used for the preparation of histological samples and their subsequent storage often compromises the integrity of DNA. A similar effect on the quality of DNA has the process happening during body decomposition. Genotyping DNA from pregnancy termination products could result in a mixed profile containing alleles from both the mother and the fetus, as it is often impossible to separate fetal and maternal tissue, especially when pregnancy is terminated at an early stage. (Semikhodskii, 2007).

DNA markers used for parentage and kinship analysis

STRs are currently the genetic markers of choice for parentage and kinship testing. In cases where an increase in discrimination power (DP) is needed to enhance the resolution of the analysis, it can be achieved by complementing the original STR marker set by other markers, e.g., single nucleotide polymorphisms (SNPs) or insertion/deletion polymorphisms (InDels) (Kayser & Knijff de, 2011; Phillips et al., 2012).

Short tandem repeat markers

STRs (microsatellites) are DNA sequences made up of a unit that usually varies in length from 2 to 7 bp and is repeated many times. Although thousands of STRs have been found in the human genome only a small subset of them is actually used for DNA identification and kinship analysis purposes (Butler, 2006). Forensic STR multiplexes commonly employ tetra-nucleotide STRs, which improve the efficiency of PCR amplification and minimize spurious DNA peaks, although penta-nucleotide STRs, like Penta E and Penta D, are also included in some STR marker panels. There are simple and complex STR loci. In simple STRs, the repeat units are constant throughout the repeat region. In contrast, complex STRs vary in their motif composition (tri, tetra-, and penta-nucleotide repeats) and may contain sequence polymorphisms within individual repeat units (Tilanus, 2006).

STR-based kinship analysis consists of generating PCR amplicons of 100−500 bp using one of several commercially available STR multiplexes, e.g., PowerPlex Fusion (Promega, USA) or VeriFiler Plus (Thermo Fisher, USA), followed by separation of the products of amplification by CE on an Applied Biosystems 3500 Genetic Analyzer (Thermo Fisher, USA) or a similar high-throughput CE instrument.

Despite being the main markers for human identification and relatedness testing, STRs suffer from several problems that complicate their use. The relatively large size

of PCR amplicons (100–500 bp) presents an issue when dealing with degraded DNA. Although this can be mitigated by employing reduced-size amplicon mini-STRs (amplicon size 60–200 bp), it is still a challenge when fixed histological preparations or samples from decomposed bodies are presented for analysis. As for noninvasive prenatal paternity testing (NIPPT) STRs have very limited use due to the nature of the circulated fetal DNA in the mother's bloodstream (see below).

Another disadvantage of STR markers is their increased mutation rate (Tracey, 2001), which often causes a mismatch between the AP and the child. Mismatches in parentage cases due to STR mutations complicate analysis and sometimes lead to false parentage exclusions. When a mutation event is suspected typing supplementary STRs or the analysis of samples from other family members may be required to correctly interpret the genotyping data.

STRs have limited information content per locus (Wenk, 2004), thus 15–20 markers or more are needed to achieve convincing levels of the CPI. Because of this, supplementary markers are often required to increase the DP, especially when a close relative is presumed to be the alternative parent or when a relationship more distant than the second degree needs to be investigated (Nothengel et al., 2010; Phillips et al., 2008). In such cases, additional STR marker panels need to be employed to find polymorphic markers to supplement the originally chosen set. This approach is costly, time-consuming, and not very efficient, as most commercially available multiplexes usually have many STRs in common and differ only in very few loci.

Single nucleotide polymorphism markers

SNPs are the most common type of genetic variation among humans. An SNP represents a difference in a single nucleotide between alleles. Most SNPs used for human identification purposes are diallelic, but recently tri-allelic SNPs have been shown to be forensically promising (Phillips et al., 2020). Compared with STRs, SNPs perform better when analyzing degraded samples (Kayser & Knijff de, 2011); they have a significantly higher number of available loci (Børsting & Morling, 2015) and a significantly lower mutation rate (Schwark et al., 2012). One major disadvantage of SNPs in parentage and kinship testing is their low discriminative power, which is offset by increasing the number of typed SNPs.

Insertion-deletion polymorphism markers

InDels are DNA length variation markers characterized by the presence or absence (insertion or deletion) of a certain DNA sequence (typically from 1 to 50 bp). They are diallelic, with great length variation between alleles, have a lower mutation rate in comparison to STRs, and are widely distributed throughout the genome, constituting about 20% of human genetic variation (Santos et al., 2010). Like SNPs, InDels have a low DP (Cortellini et al., 2020) and are mostly used as supplementary markers to STRs.

Microhaplotype markers

Microhaplotypes (MHs) are a novel type of molecular marker with a broad range of uses in forensics, parentage and kinship testing, missing person identification, and medical genetics. An MH locus is a short (less than 300 bp) segment of DNA characterized by the presence of two or more closely linked SNPs that have at least three allelic combinations (i.e., "haplotypes") (Oldoni et al., 2019). Because of the short distance between the SNPs each MH is thought to be inherited as a single block that is being passed over from generation to generation (Kidd et al., 2014).

Being multiSNP markers MHs provide, on a per locus basis, a higher DP than individual SNPs. A big advantage of MHs over STRs is that the absence of repetitive regions removes the problem of stutters caused by DNA polymerase slippage during amplification (Staadig & Tillmar, 2021).

Due to the low mutation rate of SNPs, MHs can complement standard STRs in parentage and kinship tests, especially in deficient family pedigrees (Oldoni et al., 2019), or used on their own (Staadig & Tillmar, 2021).

Genotyping technology

Capillary electrophoresis

CE is currently a standard method for STR and InDel analysis and can also be used for SNP genotyping. STR and InDel alleles are identified by the length of a PCR-amplified fragment (length-based polymorphism) and are called relative to known-sized allelic ladders, with a number value indicating the number of complete repeats observed.

Many STRs exhibit variation within the repeat region, having alleles equal in length but different in sequence. Such alleles are called isoalleles (Gettings et al., 2015). Some STR loci (e.g., D21S11, D8S1179, VWA, and several others) contain almost twice as many alleles differing in sequence than those differing in length (Gettings et al., 2015). A major disadvantage of the CE-based genotyping approach is that this technology does not allow the detection of any sequence variation between STR alleles, which may assist in making a decision as to the parental origin of the allele or even indicate a genetic mismatch. The resulting loss of information decreases the DP of the STR genotyping panel used for testing.

Another significant drawback of the CE genotyping technology is its limited multiplex capability due to the small number of dye channels available for simultaneous analysis (Grandell et al., 2016). Because amplicons of similar sizes have to be labeled with different fluorophores, the number of loci multiplexed together cannot exceed the number of dye channels in a genetic analyzer, which is 6 for the ABI 3500 family of instruments (ThermoFisher, USA) or 6-8 for the Spectrum family of instruments (Promega, USA).

The introduction of next-generation sequencing (NGS) within the framework of parentage and kinship testing provides solutions for the above limitations.

Next-generation sequencing

Second-generation sequencing, or NGS, as it is commonly known, was developed in the 1990s and became commercially available in 2005. The term "next generation" implies the next step in the development of DNA sequencing technology from the first generation (Sanger sequencing). Second-generation sequencing methods can be grouped into two major categories: sequencing by hybridization, using DNA microarrays, and sequencing by synthesis (SBS)—pyrosequencing, sequencing by ligation, ion and nanopore semiconductor-based technology (Ion Torrent, Thermo Fisher, USA, Oxford Nanopore, Oxford, UK, SBX, Stratos Genomics, Seattle, USA), and bridge amplification technology developed by Illumina (USA) (Slatko et al., 2018). All these technologies share satisfactory accuracy, high throughput, and a lower cost per base compared to Sanger sequencing (Liu et al., 2012). SBS is mainly used in medical genetics for diagnostic applications such as identifying disease-related SNPs or for molecular karyotyping (Slatko et al., 2018). NGS based on SBS has both medical and forensic applications. This is the type of NGS used for paternity and kinship testing. Bruijns et al. (2018) provide an excellent review of molecular chemistry and the technologies of various NGS platforms.

All the SBS technologies rely on high sequence coverage achieved by massive parallel sequencing (MPS), where millions of reads from a DNA fragment produced at the same time are detected on the same chip or flow cell in parallel (hence the name) (Saiz et al., 2020). Such a high number of sequencing runs is necessary to compensate for the elevated error rate of NGS relative to Sanger sequencing. After MPS, an accurate sequence of a DNA fragment is obtained as a consensus sequence of all the million reads of the same fragment (Slatko et al., 2018).

The MPS approach has several principal advantages over CE, with important implications for parentage and kinship analysis. One of the crucial features of MPS is the high level of multiplexing, both in the number of molecular markers and in the number of samples. Because MPS does not require size separation between amplicons it overcomes the multiplexing issue of current CE methods and allows typing a significantly larger number of DNA markers (not only STRs but also SNPs, InDels, MH, and mtDNA) simultaneously in a single assay (Daniel et al., 2015). In addition, many samples can be multiplexed in a single sequencing run by using barcoding, which further increases the throughput (Grandell et al., 2016). Also, because NGS requires much smaller amplicons than the CE-based STR analysis it allows obtaining genetic information from badly degraded genomic DNA.

In length-based CE genotyping, only the size of the allele is taken into consideration. In contrast, MPS could detect not only fragment length polymorphisms but also obtain detailed sequencing information about each allele analyzed (Gettings et al., 2016). This allows the discovery of previously unknown alleles and mutational events and the investigation of isoalleles and their parental origin (Rockenbauer et al., 2014), thus helping to resolve many issues affecting parentage and kinship analysis pertinent to CE. An additional

bonus of having sequence information is that SNPs and InDels are often found in the repeated units of STR loci or in the flanking regions (Rockenbauer et al., 2014), further increasing the DP of the sequence-based testing compared to conventional CE genotyping. In some cases, having sequence information may even permit differentiating homozygous twins (Weber-Lehmann et al., 2014). Because sequence information allows the differentiation of isoalleles, their origin can be determined more precisely. As a result, correct formulas for calculating relatedness statistics can be employed, increasing the accuracy of the computation.

In parentage testing, a major advantage of MPS over CE is that it permits a much deeper investigation of the nature of genetic inconsistencies. When an inconsistency is observed, CE can only determine the size of the alleles. NGS analysis goes further by obtaining information about every nucleotide in the PCR amplicon. In addition to length, it makes it possible to investigate the structure of the motif composition, sequence polymorphism within individual repeat units, and the flanking regions of the alleles. For example, a mutation or another event in the core repeat could change the size of the allele, or insertions/deletions in the flanking regions could affect primer binding (null alleles) and also the size of the allele (Silva et al., 2018).

Crucially, genotyping data obtained by MPS could be seamlessly compared to those obtained by CE-based analysis. Current CE-based allelic nomenclature conventions for STR and other length-based markers can be used to assign alleles obtained by MPS typing (Devesse et al., 2018), and the current length-based allelic frequency database could be used for calculating parentage and kinship statistics in MPS-based tests.

One major issue that prevents the wider adoption of NGS for routine parentage and kinship testing is the much higher cost per sample compared to the standard CE analysis (Alonso et al., 2018). Having said that, the increased cost of NGS analysis can be offset by the ability of this technology to deal better than CE with degraded samples, by high multiplexing, and by providing sequence information. Further development of NGS will, undoubtedly, facilitate the reduction of the cost, making it commercially competitive with the currently CE-based approach.

Commercially available NGS kits for parentage and kinship testing

A number of NGS platforms and kits for simultaneous analysis of forensically important DNA markers have been developed during the past 15 years (Saiz et al., 2020; Li et al., 2021). Currently, the Illumina MiSeq/Verogen MiSeq FGx (Illumina/Verogen, USA) and Ion GeneStudio S5 systems (ThermoFisher, USA) are the most popular instruments for forensic identity and relatedness testing. Recently, MGI developed a novel NGS sequencer, MGISEQ-2000 (MGI Tech, Co., Ltd., China), which promises to be an interesting alternative to these two systems. Available NGS platforms for forensic identification and parentage and kinship applications are discussed in detail elsewhere in this handbook.

The currently available commercial NGS multiplexes for forensic and relatedness purposes and the platforms for their analysis are presented in Table 12.1. The kits listed in the table can be loosely subdivided into three categories: (1) Kits for analysis of "standard" forensic STR markers (both autosomal and gonosmal); (2) Kits for analysis of "standard" forensic STR markers supplemented with novel or less used STRs, SNPs, and sometimes mtDNA; and (3) Kits only for analysis of SNP markers and mtDNA sequencing. Multiplexes of the first and second groups are generally used for parentage and kinship analysis, whereas those from the last group are mainly used as an auxiliary tool when more markers are needed to form an opinion as to the relationship in question.

The STR-containing multiplexes include loci that are part of the Combined DNA Index System (CODIS) and the European Standard Set, supplemented by Y-STR and X-STR, and amelogenin (Amel). Supplementary SNPs usually contain the identity and ancestry informative markers and even phenotypic informative loci. Some of the SNP-only systems have been developed to assist in specific kinship investigations. For example, the ForenSeq Kintelligence Kit targets 10,230 SNP markers and is aimed at forensic genetic genealogy research, whereas the MGIEasy Pa-SNPs Genotyping Kit is specifically designed for use in NIPPT, paternity testing, and individual identification.

Most commercially available multiplexes have been designed to be used only with a particular MPS platform. PowerSeq 46GY and ForenSeq DNA Signature Kit are compatible only with MiSeq systems (Illumina/Verogen, USA), whereas Precision ID GlobalFiler NGS STR Panel v2e and MGIEasy Signature Identification Library Prep Kit have been designed exclusively for use on Ion S5 systems (ThermoFisher Scientific, USA) and the MGISEQ-2000 (MGI Tech. Co., Ltd.), respectively. A big advantage of libraries prepared with the QIAseq family of multiplexes (Qiagen, USA) is that they can be sequenced on both MiSeq and S5 systems.

Parentage testing

Depending on when the child's sample is collected, testing for parentage can be either postnatal, when the sample is collected after the birth of the child, or prenatal when the sample is collected while the child is still in utero. The latter can be further subdivided into invasive and noninvasive depending on the type of child's biological sample used for analysis. STR markers are generally used for postnatal and invasive prenatal tests. Because of the nature of the DNA used in a child's sample SNPs are currently the markers of choice for noninvasive prenatal testing.

Postnatal parentage testing

The main issue in parentage testing is to differentiate whether an observed mismatch at the obligate allele is due to parentage exclusion or caused by mutation. Because of elevated mutation rates of STR loci (Tracey, 2001) mismatches of one or two alleles

Table 12.1 Overview of commercially available NGS kits for forensic identification, parentage, and kinship testing.

Kit name	Manufacturer	Instrument	Autosomal STR	Y-STR	X-STR	SNP	mtDNA	MH
PowerSeq 46GY[a]	Promega, USA	MiSeq	22 + Amel	23	–	–	–	–
PowerSeq CRM Nested System	Promega, USA	MiSeq	–	–	–	–	+[b]	–
ForenSeq DNA signature kit	Verogen, USA	MiSeq FGx	27 + Amel	24	7	172[c]	–	–
ForenSeq MainstAY kit	Verogen, USA	MiSeq FGx	27 + Amel	25	–	–	–	–
ForenSeq mtDNA Whole Genome kit	Verogen, USA	MiSeq FGx	–	–	–	–	+[d]	–
ForenSeq mtDNA Control Region kit	Verogen, USA	MiSeq FGx	–	–	–	–	+[b]	–
ForenSeq kintelligence kit	Verogen, USA	MiSeq FGx	–	–	–	10,230	–	–
Precision ID GlobalFiler NGS STR Panel v2[e]	ThermoFisher, USA	HID ion S5 HID ion GeneStudioS5 system	31 + Amel	3	–	–	–	–
MGIEasy signature Identification library prep kit	MGI Tech Co., China	MGISEQ-2000	51 + Amel	48	27	227[f]	+[g]	–
MGIEasy mtDNA Whole Genome Amplification kit[c]	MGI Tech Co., China	DNBSEQ-G400RS	–	–	–	–	+	–
MGIEasy Pa-SNPs Genotyping kit[h]	MGI Tech Co., China	DNBSEQ-G50 DNBSEQ-G400	–	–	–	2010	–	–
QIAseq investigator missing persons SNP panel[i]	Qiagen, USA	MiSeq ion S5 system	–	–	–	1289	–	45
QIAseq investigator ID SNP panel[j]	Qiagen, USA	MiSeq ion S5 system	–	–	–	139	–	–
QIAseq Human Mitochondria panel[d]	Qiagen, USA	MiSeq ion S5 system	–	–	–	–	+	-
QIAseq investigator microhaplotype panel	Qiagen, USA	MiSeq ion S5 system	–	–	–	–	–	+
QIAseq investigator Human Mitochondria Control Region panel[b]	Qiagen, USA	MiSeq ion S5 system	–	–	–	–	+	–

[a]The loci included are Combined DNA Index System (CODIS) core and European Standard Set (ESS) loci, amelogenin and the Y-STR loci included in the PowerPlex Y23 System.
[b]mtDNA control region.
[c]94 identity-informative SNPs, 56 ancestry-informative SNPs, and 22 phenotypic-informative SNPs.
[d]Full (16,569 bp) mitochondrial genome.
[e]Includes 21 CODIS STR markers, 9 additional multiallelic STR markers, and 4 sex determination markers.
[f]145 identity-informative SNPs, 53 ancestry-informative SNPs, and 29 phenotypic-informative SNPs.
[g]Seven amplicons covering the whole hypervariable region) of mtDNA.
[h]The kit is specifically designed for use in noninvasive prenatal paternity testing, paternity testing, and individual identification.
[i]The loci include 1200 tri-allelic kinship SNPs, 34 X-chromosome SNPs, and 55 ancestry-informative SNPs. Furthermore, the panel includes a set of 46 microhaplotypes.
[j]The panel is specifically designed to provide additional discrimination power to parentage and kinship tests.

are not uncommon in parentage testing, and nonexclusion cases with three mutations have also been described (Jacewicz et al., 2004). Information obtained by CE is often not enough to decide whether or not the genetic inconsistency is caused by mutation, and for clarification of the issue analysis of more markers and/or additional family members is required. The introduction of MPS into the framework of parentage testing makes it possible to compensate for this and other problems associated with CE-based STR typing. The sequence information obtained by MPS permits investigating the true variation within STR loci, resolves the question of whether or not an observed length-based inconsistency is due to mutation, and determines the specific paternal and maternal modes of inheritance of the allele in question.

Li and coworkers (Li et al., 2017) evaluated the performance of the Precision ID GlobalFiler NGS panel (Thermo Fisher, USA), containing 30 informative STRs and Amel, for paternity testing of a total of 6 paternity duos and one paternity trio, all with mismatched STR loci. In five cases, sequencing data from STR motifs was enough to determine that the observed genetic inconsistency was due to a mutation inherited from the father, while in the other two cases, SNP data from the flanking regions had to be additionally used to determine which allele of the AF was the source of the mutation. Compared to CE genotyping the total number of alleles observed by MPS significantly increased (176 vs. 193) as sequencing information allowed distinguishing isoalleles. These extra alleles were observed in 9 out of the 30 informative STRs. Other studies likewise report an increase in the number of STR alleles observed by MPS typing compared to the length-based analysis (Ma et al., 2016; Li et al., 2019; Silva et al., 2018). The expanded range of STR loci and a large number of SNP markers included in commercially available or custom-made MPS panels make it possible to resolve parentage without the need for repeat analysis using additional marker panels or extra family members (Ma et al., 2016).

The usefulness of MPS for tracing the origin of parental mutations was also shown in another MPS-based parentage study using the ForenSeq DNA Signature Prep Kit (Illumina, USA). Three mutations were observed in 81 parent-offspring pairs for autosomal STRs, and two mutations were observed in 26 father-son pairs for Y-STRs. By analyzing sequence information, the researchers could determine the sequence of each of these alleles and how it changed during DNA transmission for parent-offspring pairs (Li et al., 2019).

Isoalleles represent a particular issue in CE-STR parentage analysis as they exhibit no length variation between the parental and filial alleles. An isoallelic mismatch will be missed by the length-based typing, causing the wrong results to be reported. Only by using MPS the sequence variation can be detected, confirming the genetic inconsistency (Li et al., 2017; Ma et al., 2016).

The CPI values obtained by MPS genotyping are generally higher than those from the length-based analysis. In a study using ForenSeq DNA Signature Prep Kit (Illumina, USA), the average parentage index (PI) obtained by sequence-based analysis for the STR loci included in the panel (log_{10} PI = 11.52) was significantly higher than that observed

by the length-based testing for the same array of loci (log_{10} PI = 9.20). When both STR and SNP loci included in this panel were combined, the average PI values with length-based testing were several orders of magnitude higher than those observed for only sequence-based STR genotyping (log_{10} PI = 11.52 and 8.20, respectively, vs. $\log_{10}$ PI = 17.39) (Li et al., 2019).

Overall, MPS offers several key advantages over CE-STR analysis in resolving parentage cases. By capturing sequence information MPS allows the identification of a higher number of alleles than those seen by the CE-based genotyping approach. Sequence information produced by this technology permits not only differentiation between a mutation or exclusion but, crucially, determining the parental origin of the allele in question. All this, together with the ability to test large numbers of various types of DNA markers makes MPS typing more powerful for resolving parentage suites than the current CE-based methods.

The use of MPS in detecting uniparental disomy in parentage testing

MPS is an indispensable tool when parentage testing is complicated by the presence of a genetic anomaly in the child. One such anomaly that can directly affect the results of parentage testing is uniparental disomy (UPD). UPD is a rare and fascinating phenomenon manifested by the presence of a pair of homological chromosomes exceptionally derived from one parent, either from the father or from the mother (Engel, 1980). Most UPDs appear to be of no phenotypic consequences to the carrier (Kotzot, 2002), although some were found to be the reason for several important medical conditions and syndromes, such as Prader-Willy syndrome, Angelman syndrome, and transient neonatal diabetes among others. In some cases, UPDs can be the cause of homozygosity for recessive genes, leading to autosomal-recessive disorders. There are many types of UPDs: they can be constitutional or acquired, affect the whole chromosome (complete) or only a particular section (segmented), the chromosomal pair can be of paternal or maternal origin, and chromosomes derived from one parent may be identical (isodisomy) or different (heterodisomy) (Liehr, 2014).

Apart from their significance in medical genetics, UPDs often pose a problem in parentage testing, complicating distinguishing between a true exclusion and a false exclusion due to abnormal chromosomal inheritance (Mansuet-Lupo et al., 2009). The presence of UPD, especially in asymptomatic patients can be the reason for pseudoexclusion of parentage, especially in the presence of another, unrelated, mismatch (Kerr et al., 2018). The problem is particularly manifested in deficiency cases when a DNA sample of the undisputed parent is not available for analysis. Pseudoexclusions of maternity (Wegener et al., 2006) or paternity (Bein et al., 1998) due to the presence of UPD have been documented. The presence of a segmented UPD additionally complicates establishing true parentage, due to the discontinuous nature of this genetic abnormality.

When discordant results are observed and a UPD is suspected, STR length-based profiling alone is often not sufficient to confirm the presence of this chromosomal anomaly in the child. The power of MPS solves the issue in most cases. The sequence information of the alleles potentially affected by UPD can point to their parent of origin. The ability to type additional markers on the chromosome, such as SNPs, can not only provide information as to the parental origin of the chromosome (Chen et al., 2019; Su et al., 2020; Zhang et al., 2019) but also reveal whether the UPD affects the whole chromosome or only a particular fragment (Su et al., 2020).

This unusual type of chromosomal inheritance should always be considered when discordant parentage results are obtained. Regrettably, it is uncommon for parentage testing laboratories to perform additional testing, especially when the number of genetic inconsistencies observed is 1−3. Because paternity testing laboratories only in particular cases get information about a child's medical history and that this genetic abnormality is rare and predominantly asymptomatic, UPD can often be missed, and alleles affected by it are mistakenly treated as true genetic inconsistencies, causing pseudoexclusion of parentage. The problem of UPD identification is exacerbated by the relatively small number of genetic markers included in commercially available STR multiplexes and the inadequacy of the current CE genotyping technology for detecting UPDs. The power of MPS applied in parentage analysis provides the tools to further investigate the initial allelic mismatches, greatly facilitating the task of UPD identification and minimizing false parentage exclusions.

Noninvasive prenatal paternity testing

In some situations, it is necessary to perform a paternity test before the child is born. If the pregnancy is suspected to have resulted from a sexual assault or a proscribed relationship, confirmation of the identity of the biological father allows the pregnant woman to make an informed decision as to the termination or progress of the pregnancy. Until recently, paternity testing of an unborn child relied on invasive procedures such as amniocentesis, performed at gestational week (GW) 15−24, or chorionic villus sampling, performed at GW 8−12. These procedures are stressful, associated with various pregnancy-related complications, pose a serious risk of intrauterine infection and damage to the fetus, and have a chance of miscarriage of about 1% (Akolekar et al., 2015). In the early 2010s, a new, noninvasive methodology of paternity testing was developed. This approach, called NIPPT, is based on the analysis of circulating cell-free fetal DNA (ccffDNA) in maternal blood.

The ccffDNA was discovered in 1997 when Dennis Lo (Lo et al., 1997) reported the presence of cell-free fetal DNA in the blood of women during pregnancy. This type of DNA mainly originates from trophoblast cells undergoing apoptosis (Jackson, 2003). Other sources of ccffDNA are the transfer of fetal DNA through the placenta and constant placental expansion during gestation (Sekizawa et al., 2003). The relative amounts of ccffDNA vary across pregnancies and gestational week ages, reaching as high as 40%

and as reported as low as 1% (Muzzey, 2018), but generally, they range between 10% and 20% toward the latter stages of pregnancy (Lun et al., 2008). The ccffDNA first becomes detectable at GW 4–5 (Illanes et al., 2007) and disappears within 2 days *postpartum* (Lo et al., 1999). By GW 9–10 the amount of ccffDNA exceeds 3%–4% (Christiansen et al., 2019), which is sufficient for most types of noninvasive analyses, like testing for aneuploidies (Renga, 2018), monogenic disorders (Chang et al., 2021; Chen et al., 2021), blood group genotyping (Alshehri & Jackson, 2021), paternity testing (Chang et al., 2019), and so on.

The ccffDNA consists of short fragments, usually less than 160 bp in length, approximately corresponding to the length of the DNA wound around a nucleosomal core unit (Luger et al., 1997). Placenta-derived cell-free DNA fragments are shorter in length (143 bp) than maternally-derived DNA fragments (166 bp) (Renga, 2018). Although circulating cell-free DNA in the maternal bloodstream is highly fragmented, the entire fetal genome could still be determined from maternal plasma (Kitzman et al., 2012).

The small size of ccffDNA fragments has an implication on the choice of genetic markers for NIPPT. STRs are currently the standard marker system for postnatal and invasive prenatal parentage testing; however, because of the small length of ccffDNA fragments STR alleles longer than 160 bp will not be amplified. In the early attempts to perform NIPPT using the AmpFLSTR Identifiler Kit (Thermo Fisher, USA), only several informative paternally derived STR alleles were possible to detect (Wagner et al., 2009). Results obtained with the ForenSeq DNA Signature Prep Kit (Illumina, USA) were also unsatisfactory (Shen et al., 2021). In addition, the amplification of STRs often generates PCR artifacts that may be mistaken for paternally inherited fetal alleles (Christiansen et al., 2019; Shen et al., 2021). The heavy dilution of ccffDNA with maternal cell-free DNA presents further technical challenges. In this situation, genetic markers based on small amplicons, like SNPs (Chang et al., 2019; Zhang et al., 2018) and MHs (Bai et al., 2020; Ou & Qu, 2020; Wang et al., 2020), are much more informative than STRs when analyzing highly fragmented mixed DNA. Analysis of a great number of these markers by MPS allows reliably and robustly performing noninvasive paternity testing of an unborn child.

The main challenge in developing an SNP-based NIPPT is the choice and accurate genotyping of fetal SNPs in the presence of a large amount of maternal DNA. Although millions of SNPs have been identified in the human genome SNPs selected for NIPPT must be high-frequency (Chang et al., 2019; Jiang et al., 2016) and not be associated with medical conditions. To be able to detect fetally-derived SNPs, only markers for which the mother is homozygous and the fetus is heterozygous can be used. These are called "effective SNPs".

Information on the amount of fetal DNA in maternal plasma is required for correctly identifying and calling paternally inherited alleles. As only effective SNPs are used for calculating the weight of evidence about half of the SNP markers analyzed will be

uninformative in a particular study (Jiang et al., 2016; Yang et al., 2017). This has implications for the number of SNPs that need to be included in the panel to achieve the desired degree of confidence.

The GW of blood collection from the pregnant mother is also crucial for a reliable test result. Christiansen et al. (2019) investigated the chances of identifying paternally inherited alleles in maternal blood using the Precision ID Identity Panel (Thermo Fisher, USA). They have shown that at GW7, only a moderate number of paternally inherited fetal SNP alleles could be detected; at GW12, paternally inherited fetal SNP alleles were observed in the great majority of the pregnant women; and at GW20, paternally inherited fetal SNP alleles were found in all pregnant women.

Various studies have demonstrated the feasibility of detecting fetal-specific alleles in the blood of pregnant women and determining paternity. Jiang et al. (2016) developed an NIPPT based on MPS analysis of maternal plasma DNA using a custom panel containing $5x10^3$–$8x10^3$ high-frequency SNPs. Sequencing-based NIPPT results fully agreed with those obtained by invasive prenatal paternity testing using the AmpFLSTR Identifiler (Thermo Fisher, USA). In 17 out of 17 cases, paternity was determined correctly (Jiang et al., 2016).

Likewise, Chang et al. (2019) created a custom panel containing 5457 SNPs to evaluate 349 paternity cases with pregnancies of GW 6–35 and 9 negative controls from nonpregnant women who had fertility experience previously. Each mother-fetus pair was compared with all males in a blind manner. The biological fathers of all cases were correctly identified among the unrelated males. In negative controls, fetal SNPs originating from previous pregnancies could not be detected. A randomly selected subset of 50 SNPs from the total number of SNPs in the panel yielded a combined PE of 0.9999 in the full trio paternity testing (Chang et al., 2019).

In a study aimed at developing a method for NIPPT, Tam et al. (2020) employed MPS for sequencing a 356 target-SNP panel. Samples were collected at GW 7–20. Following data analysis, 108–174 target SNPs were classified as effective SNPs and included in the paternity calculations of 15 family trio cases, generating paternity probabilities of greater than 99.9999%. All paternity results were confirmed by CE-STR analysis using either amniotic fluid collected at GW 16–19 or buccal swabs collected from the newborns *postpartum*. The high specificity of the methodology was validated by successful paternity discrimination between biological fathers and their siblings and by large separations between the CPIs calculated for the biological fathers and those for 60 unrelated men.

NIPPT by MPS requires sophisticated bioinformatics to analyze large quantities of genotyping data. There is still no uniform opinion among scientists as to which biostatistical approach is correct to ensure the paternity test returns a true result. One point of view is to use the PE or PI to estimate the accuracy of an NIPPT (Drábek & Cereda, 2014; Jiang et al., 2016). The other opinion is that standard approaches based on PI are

not adequate to estimate paternity reliably. According to this school of thought conventional paternity testing methodologies rely on the availability of pure child DNA; however, the ccffDNA found in maternal plasma cannot be separated from a much larger amount of maternal cell-free DNA, thus precluding the use of traditional methods to analyze NIPPT data (Ryan et al., 2014). For this purpose, special bioinformatic algorithms that take into the account amount of the fetal fraction have been proposed for correctly identifying fetal-specific alleles and calculating the diagnostic potency of an NIPPT test (Ryan et al., 2013).

Natera Inc. (USA) was the first company to successfully commercialize the NIPPT service for the general public based on the analysis of more than 300,000 SNPs on a Human-Cyto12-SNP microarray (Ryan et al., 2013). Out of the 36,400 tests using an unrelated male as the AF, 99.95% correctly excluded paternity, and 0.05% were indeterminate. DNA Diagnostic Center developed Certainty NIPPT based on MPS analysis of 2688 SNPs (DNA Diagnostic Center, n.d) and maternal blood collection beginning from GW 10. Currently, this is the only NIPPT approved by the AABB. A number of other genetic testing providers also offer NIPPT as a service.

In contrast to many commercially available STR multiplexes for CE-based genotyping only one MPS-based panel, the MGIEasy Pa-SNPs Genotyping Kit (MGI Tech Co., China), claimed to be specifically designed for performing NIPPT by the manufacturer was commercially available at the time of writing. With the increasing number of paternity tests performed worldwide, there is little doubt that more MPS panels purposely developed for NIPPT will soon appear on the market.

Kinship testing

As it was mentioned above, kinship testing is based on evaluating the probabilities that IBS alleles between two or more individuals are IBD as opposed to being identical by chance. The more distant the relationship between the individuals, the more difficult it is to distinguish whether the relationship is true or whether the individuals are unrelated. The increase in efficiency of kinship identification can be achieved by expanding the number of genetic markers, testing additional family members, and increasing the DP of the markers analyzed.

Although modern STR multiplexes for CE analysis contain 25–30 loci, they are often inadequate to determine kinship between first-degree relatives. Establishing kinship of the second degree is even more problematic. The advent of MPS makes it possible to extend the scope of kinship testing to more distant and complex relationships.

Analysis of second-degree kinship by a custom MPS STR-only panel showed improved results when compared to CE-typing of the same loci (Liu et al., 2020). The panel contained 42 autosomal STR loci from three commercially available kits: the Goldeneye 20A System (PeopleSpot Inc., China), the Goldeneye 22NC System

(PeopleSpot Inc., China), and the Microreader 23sp System (Beijing Microread Genetics Co., Ltd., China). An increase in DP for 40 out of the 42 loci studied was observed. Family simulations and likelihood ratio (LR) calculations showed that this STR panel and ForenSeq DNA Signature Prep Kit had similar effectiveness for second-degree kinship identification.

Combining STR and SNP loci into a single sequencing panel increases the DP of kinship analysis. Li et al. (2019) evaluated the efficacy of the ForenSeq DNA Signature Prep Kit (Illumina, USA) for pairwise kinship analysis by assessing 67 individuals from two families. Six types of pairwise relationships, including 60 full siblings, 48 grandparent-grandchildren, 147 uncle/aunt-nephew/nieces, 97 first cousins, and 190 nonrelatives were generated from these two families, and the corresponding LR was calculated using either sequence-based or length-based STR genotyping data. The results showed that 54, 9, and 5 additional alleles were observed after sequencing 27 autosomal STRs, 24 Y-STRs, and 7 X-STRs, respectively, compared to those obtained by length-based information. Additionally, 11 novel alleles were identified. 51 genotypes regarded as homozygotes based on length information were found to be isoallelic heterozygotes based on sequence information. The full siblings index, grandparent-grandchild index, uncle/aunt-nephew/niece index, and first cousin index obtained by sequence-based typing were generally higher than those obtained by length-based STR analysis. When STRs and SNPs in the ForenSeq DNA Signature Prep Kit panel were combined the increase in the kinship statistics was even more pronounced. Another study using the same kit (Xu et al., 2019) and studies using SNP-only MPS panels (Grandell et al., 2016; Mo et al., 2018) confirm these observations.

Besides STRs and SNPs, MH loci also show promise in resolving kinship cases (Qu et al., 2020; Staadig & Tillmar, 2021). Staadig and Tillmar (2021) examined a custom-made QIAseq MH panel (Qiagen, USA) containing 45 different MHs by genotyping 75 individuals. The results showed that full-sibling cases could clearly be solved; however, when simulating a half-sibling versus unrelated case scenario, there was some overlap of the LR distributions, potentially resulting in inconclusiveness. The authors suggest that the inclusion of more markers in the panel would positively resolve this issue. Results of another study using a custom-designed panel of 60 MHs and 73 individuals (Qu et al., 2020) confirmed the usefulness of MPS-based analysis of MH loci for second-degree kinship inference while also noticing the need for more markers to provide more DP.

Conclusions

Since it first appeared NGS has made a deep impact in the fields of forensic identification and kinship testing. By making it possible to obtain both length-based and sequence-based information and because of its high throughput MPS technology provides strong advantages over traditional CE STR genotyping.

The sequence-based information about STR loci and their flanking regions permits the identification of new alleles and the determination of their origin. The increased allelic diversity improves the DP of the MPS analysis due to decreasing allelic frequencies relative to the number of new alleles discovered at a locus.

The MPS technology makes it easy to integrate a very large number of genetic markers into a single reaction. It has now become possible to create a genotyping panel combining autosomal and gonosomal STR, SNP, and MH loci, as well as mtDNA data. Many NGS panels are already commercially available for various purposes (Table 12.1), with the total number of loci in some of them exceeding 1000. With the increase in the number of genetic markers it might soon be possible to reliably resolve third-degree or even more distant kinship cases. Advances in bioinformatics and the development of genetic databases make it straightforward for users to design bespoke MPS panels of any complexity for a wide range of purposes.

All this makes MPS-based genotyping a powerful tool for parentage and kinship determination. At the same time, there are a number of important issues that need to be addressed prior to the wide adoption of NGS technology. Analysis of big massifs of genotyping data generated by MPS requires robust data storage and reliable bioinformatic support. The nomenclature for NGS-based STR alleles needs to be standardized, and a convention for reporting isoalleles needs to be developed. Many of the large numbers of markers included in genotyping panels are located close to each other within the same linkage group. This needs genetic linkage to be considered when developing statistical approaches for calculating parentage and kinship statistics for MPS genotyping results.

Some of these issues have already been satisfactorily resolved, and most of them, if not all, will be in the very near future. With the decrease in analysis cost, it is just a matter of short time before MPS genotyping completely replaces the current CE-based technology and becomes a method of choice for parentage and kinship analysis.

References

Akolekar, R., Beta, J., Picciarelli, G., Ogilvie, C., & D'Antonio, F. (2015). Procedure-related risk of miscarriage following amniocentesis and chorionic villus sampling: A systematic review and meta-analysis. *Ultrasound in Obstetrics and Gynecology, 45*(1), 16–26. https://doi.org/10.1002/uog.14636

Albrecht, K., & Schultheiss, D. (2004). Proof of paternity: Historical reflections on an andrological-forensic challenge. *Andrologia, 36*(1), 31–37. https://doi.org/10.1046/j.1439-0272.2003.00598.x

Alonso, A., Barrio, P. A., Müller, P., Köcher, S., Berger, B., Martin, P., Bodner, M., Willuweit, S., Parson, W., Roewer, L., & Budowle, B. (2018). Current state-of-art of STR sequencing in forensic genetics. *Electrophoresis, 39*(21), 2655–2668. https://doi.org/10.1002/elps.201800030

Alshehri, A. A., & Jackson, D. E. (2021). Non-invasive prenatal fetal blood group genotype and its application in the management of hemolytic disease of fetus and newborn: Systematic review and meta-analysis. *Transfusion Medicine Reviews, 35*(2), 85–94. https://doi.org/10.1016/j.tmrv.2021.02.001

Bai, Z., Zhao, H., Lin, S., Huang, L., He, Z., Wang, H., & Ou, X. (2020). Evaluation of a microhaplotype-based noninvasive prenatal test in twin gestations: Determination of paternity. *Zygosity and Fetal Fraction Genes (Basel), 12*(1), 26. https://doi.org/10.3390/genes12010026

Bein, G., Driller, B., Schürmann, M., Schneider, P. M., & Kirchner, H. (1998). Pseudo-exclusion from paternity due to maternal uniparental disomy 16. *International Journal of Legal Medicine, 111*, 328–330. https://doi.org/10.1007/s004140050181

Børsting, C., & Morling, N. (2015). Next generation sequencing and its applications in forensic genetics. *Forensic Science International Genetics, 18*, 78–89. https://doi.org/10.1016/j.fsigen.2015.02.002

Bruijns, B., Tiggelaar, R., & Gardeniers, H. (2018). Massively parallel sequencing techniques for forensics: A review. *Electrophoresis, 39*(21), 2642–2654. https://doi.org/10.1002/elps.201800082

Butler, J. M. (2006). Genetics and genomics of core short tandem repeat loci used in human identity testing. *Journal of Forensic Sciences, 51*(2), 253–265. https://doi.org/10.1111/j.1556-4029.2006.00046.x

Chakraborty, R., Shaw, M., & Schull, W. J. (1974). Exclusion of paternity: The current state of the art. *The American Journal of Human Genetics, 26*(4), 477–488.

Chang, L., Yu, H., Miao, X., Zhang, J., & Li, S. (2019). Development and comprehensive evaluation of a noninvasive prenatal paternity testing method through a scaled trial. *Forensic Science International Genetics, 43*, 102158. https://doi.org/10.1016/j.fsigen.2019.102158

Chang, L., Zhu, X., Li, R., Wu, H., Chen, W., Chen, J., Liu, H., Li, S., & Liu, P. (2021). A novel method for noninvasive diagnosis of monogenic diseases from circulating fetal cells. *Prenatal Diagnosis, 41*(4), 400–408. https://doi.org/10.1002/pd.5796

Chen, M., Jiang, J., Li, C., Ren, H., Chen, W., Liu, Z., Cheng, F., Zhao, J., Chen, T., Chen, C., & Yan, J. (2019). Non-pathological complete paternal uniparental isodisomy of chromosome 2 revealed in a maternity testing case. *International Journal of Legal Medicine, 133*(4), 993–997. https://doi.org/10.1007/s00414-018-1857-x

Chen, C., Li, R., Sun, J., Zhu, Y., Jiang, L., Li, J., Fu, F., Wan, J., Guo, F., An, X., Wang, Y., Fan, L., Sun, Y., Guo, X., Zhao, S., Wang, W., Zeng, F., Yang, Y., Ni, P., … Liao, C. (2021). Noninvasive prenatal testing of α-thalassemia and β-thalassemia through population-based parental haplotyping. *Genome Medicine, 13*(1), 18. https://doi.org/10.1186/s13073-021-00836-8

Christiansen, S. L., Jakobsen, B., Børsting, C., Udengaard, H., Buchard, A., Kampmann, M. L., Grøndahl, M. L., & Morling, N. (2019). Non-invasive prenatal paternity testing using a standard forensic genetic massively parallel sequencing assay for amplification of human identification SNPs. *International Journal of Legal Medicine, 133*(5), 1361–1368. https://doi.org/10.1007/s00414-019-02106-0

Cortellini, V., Brescia, G., Carnevali, E., Cerri, N., Correa, H. S. D., Nespeca, P., Severini, S., Tommolini, F., Tozzo, P., Verzeletti, A., & Caenazzo, L. (2020). Genetic data and comparative study of 38 autosomal InDel markers in three Italian population groups. *Forensic Science International Genetics, 44*, 102170. https://doi.org/10.1016/j.fsigen.2019.102170

Daniel, R., Santos, C., Phillips, C., Fondevila, M., van Oorschot, R. A., Carracedo, A., Lareu, M. V., & McNevin, D. (2015). A SNaPshot of next generation sequencing for forensic SNP analysis. *Forensic Science International Genetics, 14*, 50–60. https://doi.org/10.1016/j.fsigen.2014.08.013

Devesse, L., Ballard, D., Davenport, L., Riethorst, I., Mason-Buck, G., & Syndercombe Court, D. (2018). Concordance of the ForenSeq system and characterisation of sequence-specific autosomal STR alleles across two major population groups. *Forensic Science International Genetics, 34*, 57–61. https://doi.org/10.1016/j.fsigen.2017.10.012

DNA Diagnostic Center https://dnacenter.com/paternity-testing/non-invasive-prenatal-paternity-testing/ Accessed 21 November 2021.

Drábek, J., & Cereda, G. (2014). Interpreting noninvasive prenatal paternity tests. *Genetics in Medicine, 16*(10), 793–794. https://doi.org/10.1038/gim.2014.100

Dungern, E von, & Hirschfeld, L. (1910). Ueber Vererbung 5 gruppenspezifischer strukturen des blutes. *Z Immunforsch, 6*, 284–292.

Engel, E. (1980). A new genetic concept: Uniparental disomy and its potential effect, isodisomy. *American Journal Medicine Genetics, 6*, 137–143. https://doi.org/10.1002/ajmg.1320060207

Gettings, K. B., Aponte, R. A., Vallone, P. M., & Butler, J. M. (2015). STR allele sequence variation: Current knowledge and future issues. *Forensic Science International Genetics, 18*, 118–130.

Gettings, K. B., Kiesler, K. M., Faith, S. A., Montano, E., Baker, C. H., Young, B. A., Guerrieri, R. A., & Vallone, P. M. (2016). Sequence variation of 22 autosomal STR loci detected by next generation sequencing. *Forensic Science International Genetics, 21*, 15–21.

Gill, P., Jeffreys, A., & Werrett, D. (1985). Forensic application of DNA 'fingerprints. *Nature, 318*, 577–579. https://doi.org/10.1038/318577a0

Gjertson, D. W., Brenner, C. H., Baur, M. P., Carracedo, A., Guidet, F., Luque, J. A., Lessig, R., Mayr, W. R., Pascali, V. L., Prinz, M., Schneider, P. M., & Morling, N. (2007). ISFG: Recommendations on biostatistics in paternity testing. *Forensic Science International Genetics, 1*(3–4), 223–231. https://doi.org/10.1016/j.fsigen.2007.06.006

Grandell, I., Samara, R., & Tillmar, A. O. (2016). A SNP panel for identity and kinship testing using massive parallel sequencing. *International Journal of Legal Medicine, 130*(4), 905–914. https://doi.org/10.1007/s00414-016-1341-4

Illanes, S., Denbow, M., Kailasam, C., Finning, K., & Soothill, P. W. (2007). Early detection of cell-f ree fetal DNA in maternal plasma. *Early Human Development, 83*(9), 563–566. https://doi.org/10.1016/j.earlhumdev.2006.11.001

Jacewicz, R., Berent, J., Prosniak, A., Dobosz, T., Kowalczyk, E., & Szram, S. (2004). Non-exclusion paternity case with a triple genetic incompatibility. *International Congress Series, 1261*, 511–513.

Jackson, L. (2003). Fetal cells and DNA in maternal blood. *Prenatal Diagnosis, 23*(10), 837–846. https://doi.org/10.1002/pd.705

Jeffreys, A., Brookfield, J., & Semeonoff, R. (1985). Positive identification of an immigration test-case using human DNA fingerprints. *Nature, 317*, 818–819. https://doi.org/10.1038/317818a0

Jiang, H., Xie, Y., Li, X., Ge, H., Deng, Y., Mu, H., Feng, X., Yin, L., Du, Z., Chen, F., & He, N. (2016). Noninvasive prenatal paternity testing (nipat) through maternal plasma DNA sequencing: A pilot study. *PLoS One, 11*(9), e0159385. https://doi.org/10.1371/journal.pone.0159385

Kayser, M., & Knijff de, P. (2011). Improving human forensics through advances in genetics, genomics and molecular biology. *Nature Reviews Genetics, 12*, 179–192. https://doi.org/10.1038/nrg2952

Kerr, E. R., Stuhlmiller, G. M., Maha, G. C., Ladd, M. A., Mikhail, F. M., Koester, R. P., & Hurst, A. C. E. (2018). Maternal uniparental isodisomy for chromosome 6 discovered by paternity testing: A case report. *Molecular Cytogenetics, 11*, 60. https://doi.org/10.1186/s13039-018-0411-3

Kidd, K. K., Pakstis, A. J., Speed, W. C., Lagacé, R., Chang, J., Wootton, S., Haigh, E., & Kidd, J. R. (2014). Current sequencing technology makes microhaplotypes a powerful new type of genetic marker for forensics. *Forensic Science International Genetics, 12*, 215–224. https://doi.org/10.1016/j.fsigen.2014.06.014

Kitzman, J. O., Snyder, M. W., Ventura, M., Lewis, A. P., Qiu, R., Simmons, L. E., Gammill, H. S., Rubens, C. E., Santillan, D. A., Murray, J. C., Tabor, H. K., Bamshad, M. J., Eichler, E. E., & Shendure, J. (2012). Noninvasive whole-genome sequencing of a human fetus. *Science Translational Medicine, 4*(137), 137ra76. https://doi.org/10.1126/scitranslmed.3004323

Kotzot, D. (2002). Review and meta-analysis of systematic searches for uniparental disomy (UPD) other than UPD 15. *American Journal Medicine Genetics, 111*, 366–375. https://doi.org/10.1002/ajmg.10569

Landsteiner, K. (1901). Ueber agglutinationserscheinungen normalen menschlichen blutes. *Wiener Klinische Wochenschrift, 14*, 1132–1134.

Liehr, T. (2014). *Uniparental disomy (UPD) in clinical genetics*. Berlin, Heidelberg: Springer-Verlag.

Li, R., Li, H., Peng, D., Hao, B., Wang, Z., Huang, E., Wu, R., & Sun, H. (2019). Improved pairwise kinship analysis using massively parallel sequencing. *Forensic Science International Genetics, 38*, 77–85. https://doi.org/10.1016/j.fsigen.2018.10.006

Li, R., Shen, X., Chen, H., Peng, D., Wu, R., & Sun, H. (2021). Developmental validation of the MGIEasy signature identification library prep kit, an all-in-one multiplex system for forensic applications. *International Journal of Legal Medicine, 135*(3), 739–753. https://doi.org/10.1007/s00414-021-02507-0

Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., Lin, D., Lu, L., & Law, M. (2012). Comparison of next-generation sequencing systems. *Journal of Biomedicine and Biotechnology, 2012*, 251364. https://doi.org/10.1155/2012/251364

Liu, Q., Ma, G., Du, Q., Lu, C., Fu, L., Wang, Q., Fu, G., Li, S., & Cong, B. (2020). Development of an NGS panel containing 42 autosomal STR loci and the evaluation focusing on secondary kinship analysis. *International Journal of Legal Medicine, 134*(6), 2005–2014. https://doi.org/10.1007/s00414-020-02295-z

Li, H., Zhao, X., Ma, K., Cao, Y., Zhou, H., Ping, Y., Shao, C., Xie, J., & Liu, W. (2017). Applying massively parallel sequencing to paternity testing on the Ion torrent personal genome machine. *Forensic Science International Genetics, 31*, 155–159. https://doi.org/10.1016/j.fsigen.2017.09.007

Lo, Y. M., Corbetta, N., Chamberlain, P. F., Rai, V., Sargent, I. L., Redman, C. W., & Wainscoat, J. S. (1997). Presence of fetal DNA in maternal plasma and serum. *Lancet, 350*(9076), 485–487. https://doi.org/10.1016/S0140-6736(97)02174-0

Lo, Y. M., Zhang, J., Leung, T. N., Lau, T. K., Chang, A. M., & Hjelm, N. M. (1999). Rapid clearance of fetal DNA from maternal plasma. *The American Journal of Human Genetics, 64*(1), 218–224. https://doi.org/10.1086/302205

Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F., & Richmond, T. J. (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature, 389*(6648), 251–260. https://doi.org/10.1038/38444

Lun, F. M., Chiu, R. W., Chan, K. C., Leung, T. Y., Lau, T. K., & Lo, Y. M. (2008). Microfluidics digital PCR reveals a higher than expected fraction of fetal DNA in maternal plasma. *Clinical Chemistry, 54*(10), 1664–1672. https://doi.org/10.1373/clinchem.2008.111385

Ma, Y., Kuang, J. Z., Nie, T. G., Zhu, W., & Yang, Z. (2016). Next generation sequencing: Improved resolution for paternal/maternal duos analysis. *Forensic Science International Genetics, 24*, 83–85. https://doi.org/10.1016/j.fsigen.2016.05.015

Mansuet-Lupo, A., Henke, J., Henke, L., Blank, C., Ernsting, A., Kozlowski, P., Rouger, P., & Van Huffel, V. (2009). A paternity case with three genetic incompatibilities between father and child due to maternal uniparental disomy 21 and a mutation at the Y chromosome. *Forensic Science International Genetics, 3*(2), 141–143. https://doi.org/10.1016/j.fsigen.2008.09.010

Mo, S. K., Ren, Z. L., Yang, Y. R., Liu, Y. C., Zhang, J. J., Wu, H. J., Li, Z., Bo, X. C., Wang, S. Q., Yan, J. W., & Ni, M. (2018). A 472-SNP panel for pairwise kinship testing of second-degree relatives. *Forensic Science International Genetics, 34*, 178–185. https://doi.org/10.1016/j.fsigen.2018.02.019

Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G., & Erlich, H. (1986). Specific enzymatic amplification of DNA in vitro: The polymerase chain reaction. *Cold Spring Harbor Symposia on Quantitative Biology, 51*(Pt 1), 263–273. https://doi.org/10.1101/sqb.1986.051.01.032

Muzzey, D. (2018). The technology and bioinformatics of cell-free DNA-based NIPT. In L. Page-Christiaens, & H.-G. Klein (Eds.), *Noninvasive prenatal testing (NIPT). Applied genomics in prenatal screening and diagnosis*. Academic Press, 2018.

Nothnagel, M., Schmidtke, J., & Krawczak, M. (2010). Potentials and limits of pairwise kinship analysis using autosomal short tandem repeat loci. *International Journal of Legal Medicine, 124*(3), 205–215. https://doi.org/10.1007/s00414-009-0413-0

Oldoni, F., Kidd, K. K., & Podini, D. (2019). Microhaplotypes in forensic genetics. *Forensic Science International Genetics, 38*, 54–69. https://doi.org/10.1016/j.fsigen.2018.09.009

Ottenberg, R. (1921). Medicolegal application of human blood grouping. *JAMA, 77*, 682–683.

Ou, X., & Qu, N. (2020). Noninvasive prenatal paternity testing by target sequencing microhaps. *Forensic Science International Genetics, 48*, 102338. https://doi.org/10.1016/j.fsigen.2020.102338

Phillips, C., Amigo, J., Tillmar, A. O., Peck, M. A., de la Puente, M., Ruiz-Ramírez, J., Bittner, F., Idrizbegović, Š., Wang, Y., Parsons, T. J., & Lareu, M. V. (2020). A compilation of tri-allelic SNPs from 1000 Genomes and use of the most polymorphic loci for a large-scale human identification panel. *Forensic Science International Genetics, 46*, 102232. https://doi.org/10.1016/j.fsigen.2020.102232

Phillips, C., Fondevila, M., Garcia-Magarinos, M., Rodriguez, A., Salas, A., Carracedo, A., & Lareu, M. V. (2008). Resolving relationship tests that show ambiguous STR results using autosomal SNPs as supplementary markers. *Forensic Science International Genetics, 2*(3), 198–204. https://doi.org/10.1016/j.fsigen.2008.02.002

Phillips, C., García-Magariños, M., Salas, A., Carracedo, A., & Lareu, M. V. (2012). SNPs as supplements in simple kinship analysis or as core markers in distant pairwise relationship tests: When do SNPs add value or replace well-established and powerful STR tests? *Transfusion Medicine and Hemotherapy, 39*(3), 202–210. https://doi.org/10.1159/000338857

Qu, N., Lin, S., Gao, Y., Liang, H., Zhao, H., & Ou, X. (2020). A microhap panel for kinship analysis through massively parallel sequencing technology. *Electrophoresis, 41*(3–4), 246–253. https://doi.org/10.1002/elps.201900337

Renga, B. (2018). Non invasive prenatal diagnosis of fetal aneuploidy using cell free fetal DNA. *European Journal of Obstetrics & Gynecology and Reproductive Biology, 225*, 5–8. https://doi.org/10.1016/j.ejogrb.2018.03.033

Rockenbauer, E., Hansen, S., Mikkelsen, M., Børsting, C., & Morling, N. (2014). Characterization of mutations and sequence variants in the D21S11 locus by next generation sequencing. *Forensic Science International Genetics, 8*, 68–72.

Ryan, A., Baner, J., Demko, Z., Hill, M., Sigurjonsson, S., Baird, M. L., & Rabinowitz, M. (2013). Informatics-based, highly accurate, noninvasive prenatal paternity testing. *Genetics in Medicine, 15*(6), 473–477. https://doi.org/10.1038/gim.2012.155

Ryan, A., Demko, Z., Sigurjonsson, S., & Rabinowitz, M. (2014). Response to Drábek and Cereda. *Genetics in Medicine, 16*(10), 794. https://doi.org/10.1038/gim.2014.99

Saiz, M., Martínez-Gonzalez, L. J., Lorente, J. A., & Alvarez-Cubero, M. J. (2020). Applications of next-generation sequencing in forensic field. In P. Shrivastava, H. R. Dash, J. A. Lorente, & J. Imam (Eds.), *Forensic DNA typing: Principles, applications and advancements*. Singapore: Springer. https://doi.org/10.1007/978-981-15-6655-4_23

Santos, N. P., Ribeiro-Rodrigues, E. M., Ribeiro-Dos-Santos, A. K., Pereira, R., Gusmão, L., Amorim, A., Guerreiro, J. F., Zago, M. A., Matte, C., Hutz, M. H., & Santos, S. E. (2010). Assessing individual interethnic admixture and population substructure using a 48-insertion-deletion (INSEL) ancestry-informative marker (AIM) panel. *Human Mutation, 31*(2), 184–190. https://doi.org/10.1002/humu.21159

Schwark, T., Meyer, P., Harder, M., Modrow, J. H., & von Wurmb-Schwark, N. (2012). The SNPforID assay as a supplementary method in kinship and trace analysis. *Transfusion Medicine and Hemotherapy, 39*(3), 187–193. https://doi.org/10.1159/000338855

Sekizawa, A., Yokokawa, K., Sugito, Y., Iwasaki, M., Yukimoto, Y., Ichizuka, K., Saito, H., & Okai, T. (2003). Evaluation of bidirectional transfer of plasma DNA through placenta. *Human Genetics, 113*(4), 307–310. https://doi.org/10.1007/s00439-003-0987-4

Semikhodskii, A. (2007). *Dealing with DNA evidence: A legal guide*. London: Routledge Cavendish.

Shaw, M. W. (1983). Paternity determination: 1921 to 1983 and beyond. *JAMA, 250*(18), 2536–2537. https://doi.org/10.1001/jama.1983.03340180090038

Shen, X., Li, R., Li, H., Gao, Y., Chen, H., Qu, N., Peng, D., Wu, R., & Sun, H. (2021). Noninvasive prenatal paternity testing with a combination of well-established SNP and STR markers using massively parallel sequencing. *Genes, 12*(3), 454. https://doi.org/10.3390/genes12030454

Silva, D. S. B. S., Sawitzki, F. R., Scheible, M. K. R., Bailey, S. F., S Alho, C., & Faith, S. A. (2018). Paternity testing using massively parallel sequencing and the PowerSeq AUTO/Y system for short tandem repeat sequencing. *Electrophoresis, 39*(21), 2669–2673. https://doi.org/10.1002/elps.201800072

Slatko, B. E., Gardner, A. F., & Ausubel, F. M. (2018). Overview of next-generation sequencing technologies. *Current Protocols in Molecular Biology, 122*(1), e59. https://doi.org/10.1002/cpmb.59

Staadig, A., & Tillmar, A. (2021). Evaluation of microhaplotypes in forensic kinship analysis from a Swedish population perspective. *International Journal of Legal Medicine, 135*(4), 1151–1160. https://doi.org/10.1007/s00414-021-02509-y

Su, H., Sun, T., Chen, M., Liu, J., Wang, X., Chen, Y., Ren, W., Zhang, G., Yan, J., & Yun, K. (2020). Multiple methods used for type detection of uniparental disomy in paternity testing. *International Journal of Legal Medicine, 134*(3), 885–893. https://doi.org/10.1007/s00414-019-02215-w

Tam, J. C. W., Chan, Y. M., Tsang, S. Y., Yau, C. I., Yeung, S. Y., Au, K. K., & Chow, C. K. (2020). Noninvasive prenatal paternity testing by means of SNP-based targeted sequencing. *Prenatal Diagnosis, 40*(4), 497–506. https://doi.org/10.1002/pd.5595

The Relationship Testing Standards Program Unit. (2013). *Annual report summary*. *Association for the advancement of blood and Biotherapies website*. Available at: https://www.aabb.org/docs/default-source/default-document-library/accreditation/2013-relationship-testing-summary-report.pdf. (Accessed 19 December 2021).

Tilanus, M. G. (2006). Short tandem repeat markers in diagnostics: what's in a repeat? *Leukemia, 20*(8), 1353–1355. https://doi.org/10.1038/sj.leu.2404273

Tracey, M. (2001). Short tandem repeat-based identification of individuals and parents. *Croatian Medical Journal, 42*(3), 233–238.

Wagner, J., Dzijan, S., Marjanović, D., & Lauc, G. (2009). Non-invasive prenatal paternity testing from maternal blood. *International Journal of Legal Medicine, 123*(1), 75–79. https://doi.org/10.1007/s00414-008-0292-9

Wang, J. Y. T., Whittle, M. R., Puga, R. D., Yambartsev, A., Fujita, A., & Nakaya, H. I. (2020). Noninvasive prenatal paternity determination using microhaplotypes: A pilot study. *BMC Medical Genomics, 13*(1), 157. https://doi.org/10.1186/s12920-020-00806-w

Weber-Lehmann, J., Schilling, E., Gradl, G., Richter, D. C., Wiehler, J., & Rolf, B. (2014). Finding the needle in the haystack: Differentiating "identical" twins in paternity testing and forensics by ultra-deep next generation sequencing. *Forensic Science International Genetics, 9*, 42–46. https://doi.org/10.1016/j.fsigen.2013.10.015

Wegener, R., Weirich, V., Dauber, E. M., & Mayr, W. R. (2006). Mother-child exclusion due to paternal uniparental disomy 6. *International Journal of Legal Medicine, 120*, 282–285. https://doi.org/10.1007/s00414-006-0077-y

Wenk, R. E. (2004). Testing for parentage and kinship. *Current Opinion in Hematology, 11*(5), 357–361. https://doi.org/10.1097/01.moh.0000137914.80855.8a

Xu, M., Du, Q., Ma, G., Chen, Z., Liu, Q., Fu, L., Cong, B., & Li, S. (2019). Utility of ForenSeq DNA signature Prep kit in the research of pairwise 2nd-degree kinship identification. *International Journal of Legal Medicine, 133*(6), 1641–1650. https://doi.org/10.1007/s00414-019-02003-6

Yang, D., Liang, H., Gao, Y., Lin, S., He, Z., Gao, J., Sun, H., Li, Q., Ma, X., & Ou, X. (2017). Noninvasive fetal genotyping of paternally inherited alleles using targeted massively parallel sequencing in parentage testing cases. *Transfusion, 57*(6), 1505–1514. https://doi.org/10.1111/trf.14077

Zhang, X., Ding, Z., He, R., Qi, J., Zhang, Z., & Cui, B. (2019). Complete paternal uniparental disomy of chromosome 2 in an asian female identified by short tandem repeats and whole genome sequencing. *Cytogenetic and Genome Research, 157*(4), 197–202. https://doi.org/10.1159/000499893

Zhang, S., Han, S., Zhang, M., & Wang, Y. (2018). Non-invasive prenatal paternity testing using cell-free fetal DNA from maternal plasma: DNA isolation and genetic marker studies. *Legal Medicine, 32*, 98–103. https://doi.org/10.1016/j.legalmed.2018.03.009

This page intentionally left blank

CHAPTER 13

Forensic relevance of SNP analysis in next-generation sequencing

Seema Malhotra and Shivani Sehgal
School of Forensic Sciences, National Forensic Sciences University (NFSU), Delhi Campus, Delhi, India

Introduction

Forensic genetic analysis is a powerful tool in the identification of criminals based on various biological forensic evidence such as blood, semen, saliva, hair, bone, teeth, and muscle tissues. The first method of routine human identity testing was developed in the mid-1980s in the United States, called restriction fragment length polymorphism (RFLP) (Budowle and Baechtel, 1990; Jeffreys et al., 1985; Wyman & White, 1980). Typical VNTR loci having high power of discrimination were typed successfully using RFLP analysis, with a minimum DNA requirement of at least 10–25 ng (Giusti & Budowle, 1995). Because of the large amount of DNA template requirement, the RFLP technique was soon replaced with polymerase chain reaction (PCR)-dependent methods for typing the genetic markers (Saiki et al., 1985). Very low amounts of DNA up to picogram level were routinely typed with the use of PCR in forensic casework over the last few decades.

The initial genetic marker system utilizing PCR-based assays was dependent on single nucleotide polymorphism (SNP) variations. The two loci named HLA-DQA1 and PolyMarker loci consisting of sequence-based polymorphism were analyzed using ASO (Allele-specific oligonucleotide) hybridization probes in a reverse dot blot format (Budowle et al., 1995; Comey and Budowle, 1991; Gyllensten and Erlich, 1988; Walsh et al., 1991). Further, these SNP-based systems despite providing highly sensitive detection were discouraged because of their limited power of discrimination. This led to the shift from early PCR-based SNP genetic markers to short tandem repeat (STR) markers, or microsatellites, which have now become an essential part of forensic DNA genotyping. Multiplexing, small product size, polymorphic nature, and easy amplification by PCR are the key parameters of an STR marker (Budowle et al., 2001). Consequently, the currently employed multiplex autosomal STR loci for DNA typing of forensic samples offer high sensitivity, specificity, and power of discrimination with the ability to type low DNA template samples. However, in cases of highly degraded biological samples with limited DNA, SNPs have proven to be better than STR typing. Thus, for

Next Generation Sequencing (NGS) Technology in DNA Analysis
ISBN 978-0-323-99144-5, https://doi.org/10.1016/B978-0-323-99144-5.00015-9
© 2024 Elsevier Inc.
All rights reserved.

generating profiles from degraded samples and for the higher yield of genetic information from forensically compromised samples, SNPs serve as a potential alternative to STRs (Budowle et al., 2004).

Characterization and origin of SNPs

The haploid human genome consisting of approximately 3.2 billion nucleotides, exhibits both length- and sequence-based genetic variations (Batnaym et al., 2013). Length-based genetic variations represented by STRs are made up of tandem repeated sequences, of which each is 2–7 base pairs in length. Sequence variants representing SNPs are variations at a single base arising due to point mutations at definite locations in the genome. Mutations like base substitutions, either transitions or trans versions, give rise to two or more alternative bases at a given position. Any Genetic variability at a base position is treated as an SNP only when a minimum of two alleles have frequencies greater than 1% in a population. The number of SNPs is calculated to be 10–11 million in the human genome, showing the presence of one SNP per 275 bp in the human genome (Sachidanandam et al., 2001; Venter et al., 2001). Although multiallelic SNPs do exist, the majority of the SNPs are usually biallelic because of the low recurrence of mutation at one nucleotide, and thus it is highly unlikely that two independent base changes occur at a single position over time. Therefore, the frequency of single nucleotide substitution in the mammalian genome is estimated to be in the range of 1×10^{-9} and 5×10^{-9} per nucleotide per year (Martínez-Arias et al., 2001). Since 85% of human genetic variations are due to SNPs, they are ubiquitous and the most common class of human polymorphism (Holden, 2002). Some of the SNPs are restricted to one major population and are only termed as rare variants or single nucleotide variants. However, most of the SNPs are polymorphic in all human populations and are called true SNPs.

Depending upon origin, SNPs can be found in both coding and noncoding sites of the genome. When they are found in the coding sites of the genes, they cause variation in the structure or function of the encoded protein and become an important tool in the diagnosis of genetic disorders. Such an SNP that affects a protein is termed a coding SNP (cSNP) or an exonic SNP. However, if they are found in the noncoding sites of the genome, the phenotype of an individual is not affected; rather, these are used in evolutionary studies, population genetics, and forensic analysis (Gill, 2001; Zhao et al., 2003). The application of SNP markers is well established in biological science, as whole human genome sequencing resulted in the finding of millions of SNPs, which are being applied in various areas of forensic science, i.e., phenotype prediction, testing of mitochondrial DNA (mtDNA), Y-SNPs as lineage markers and ancestry informative markers, and other probable forensic casework applications. SNPs hold potential future roles with the aim of developing an ideal typing method having a high level of precision and sensitivity adequate for the analysis of degraded DNA along with good multiplexing capacity.

Conventional SNP genotyping assays

Over the last few years, multiple SNP analysis techniques have been observed. The conventional methods of SNP typing occur on the basis of four major molecular mechanisms, i.e., allele-specific hybridization, primer extension, oligonucleotide ligation, and invasive cleavage. A summary of these four major molecular mechanisms for allelic discrimination along with several methods of detection for analyzing the products is given in Table 13.1. However, an ideal SNP genotyping technique meeting all research requirements, particularly for forensic purposes where a low amount of DNA is available for casework, is still lacking. Some applications work with little DNA quantity but require a large number of SNPs, whereas others use few SNPs but a large quantity of DNA. Besides, the use of both SNPs and input DNA in large quantities thus results in varying levels of throughput depending upon the application used.

Moreover, the requirement of PCR amplification in almost all conventional methods limits its use in forensic applications where multiplexing is being carried out. Techniques like the Invader Assay analyze SNP markers without carrying out a PCR process, but the requirement of a large amount of DNA makes them unsuitable for forensic analysis. Therefore, conventional SNP genotyping techniques with limited multiplexing capability are not suitable for routine forensic DNA analysis. Another limitation lies in the mixture analysis of a sample due to the bi-allelic behavior of the majority of SNP markers. However, quantification techniques like mass spectrometry and pyrosequencing have enabled the determination of contributors in a mixed profile. The various detection platforms that have existed for a long time for SNP typing have prevented its routine analysis because of the lack of adoption of one gold standard method for SNP typing (Budowle, 2004; Sobrino et al., 2005). Hence, while choosing a technique for SNP typing, the parameters that need to be considered are accuracy, cost, sensitivity, flexibility, and time consumption.

Benefits of next generation sequencing technology over conventional genotyping assays

Overview

The determination of the nucleotides in a polymer of nucleic acid is the major goal of sequencing. First Sequencing was conducted with the invention of the Sanger sequencing method in 1977, where the principle of sequencing by synthesis (SBS) method allowed the incorporation of a ddNTP into a growing DNA chain, thereby terminating further extension (Sanger et al., 1977). With the subsequent introduction of fluorescently labeled ddNTPs and capillary electrophoresis (CE) platforms, the sequencing of complete genomes became possible (Mitchelson, 2003). Over the last decade, advanced sequencing technologies have surpassed Sanger-based sequencing in

Table 13.1 Summary of major SNP genotyping assays along with principle and detection method for each allelic discrimination reaction.

S.no	SNP typing assay methods	Principle	Detection method	Methodology	Advantages	Disadvantages	References
1	**Allele-specific hybridization/ ASO (allele-specific oligonucleotide hybridization)**	Using ASO probes DNA targets varying at one nucleotide are discriminated by hybridization.	1. Fluorescence resonance energy transfer (FRET) (homogenous hybridization)	Two PCR primers along with fluorescently labelled ASO probes. Increased amount of DNA corresponds to increased fluorescence.	Highly sensitive. High throughput genotyping. Single tube reaction for PCR and detection.	Limited multiplexing capability.	Wallace et al. (1979), Clegg (1992)
			1.1 Using LightCycler (ROCHE), which measures the light intensity.	One probe labelled with fluorescein at its 3′ end and another with LC red 640 or 705 label at its 5′ end are hybridized on the DNA target.			Lareu et al. (2001)
			1.2 Using TAQMAN assay (Applied Biosystems) (5′ nuclease activity of *Taq* polymerase displaces and cleaves probes hybridized to the target).	Two TaqMan probes with fluorescent dyes at 5′ end and a quencher at 3′ end hybridize to the target DNA, with cleaving of 5′ dye by 5′nuclease activity of *Taq* polymerase.			Holland et al. (1991)

			1.3 Using molecular beacons (oligonucleotide probes with two complementary sequences flanking the complementary sequence to the target, with a fluorophore at the 5′ end and a quencher at its 3′ end).	When two molecular beacons with different fluorophores when binding to the target DNA, the quencher and the fluorophore are separated, and the fluorescence appears.			Tyagi and Kramer (1996)
			2. Microarrays (array hybridization) e.g., GeneChip systems (affymetrix) for each SNP use tens of ASO probes.	Short oligonucleotides bind to solid supports hybridized with fluorescently labelled PCR products.	GeneChip is designed for genotyping a large number of SNPs.	Laborious to create ideal conditions for multiplexing.	Mei et al. (2000)
2	**Primer extension**	Based on the capability of DNA polymerase to introduce specific dNTPs complementary to the template DNA sequence. Three types of reactions exist:					Kuppuswamy et al. (1991)

Continued

Table 13.1 Summary of major SNP genotyping assays along with principle and detection method for each allelic discrimination reaction.—cont'd

S.no	SNP typing assay methods	Principle	Detection method	Methodology	Advantages	Disadvantages	References
		1. Minisequencing reaction or single nucleotide primer extension. DNA polymerase performs extension of a primer that anneals to its target DNA immediately adjacent to the SNP.	1.1 SNaPshot (Applied Biosystems) uses electrophoresis.	An unlabelled primer at 3′ end upstream to the SNP region is extended with one fluorescently labelled ddNTP.	Multiplex reactions can be done by structural separation of the mini-sequencing products.		Quintáns et al. (2004), Sanchez et al. (2005)
			1.2 MALDI-TOF MS (matrix assisted laser desorption ionization time of flight Mass spectrometer) e.g., MassEXTEND, sequenom, PinPoint assay (Applied biosystems).	Determines the molecular weight of a DNA product by calculating the time of flight.	Resolution is very high. Direct method of detection.	Difficult to differentiate between heterozygous and homozygous genotypes. Pure sample required.	Haff & Smirnov (1997), Braun et al. (1997)
			1.3 Microarrays (fluorescence)	Performed on the surface of the chip or in solution for genotyping SNPs with mini-sequencing primers.	An array of arrays format allows analysis of many samples per slide.	Difficult to design.	Pastinen et al. (2000)

			1.4 Fluorescence polarization (FP)	Primer extended by the allele specific dye-terminator on the template, thereby increasing the molecular weight of the fluorophore.	Used in detecting allelic discrimination products where the initial molecules have different sizes.		Chen et al. (1999)
		2. Allele-specific extension reaction works on the difference in extension efficiency of DNA polymerase between primers with matched and mismatched 3′ ends.	2.1 Microarrays (fluorescence)	Two primers are needed, one for each SNP allele. Primer forming the product is detected using fluorescently labelled nucleotides on a microarray.			Pastinen et al. (2000)
		3. Pyrosequencing reaction based on sequencing by synthesis method.	3.1 Luminescence (detection based on the pyrophosphate released during the DNA polymerase reaction)	Light is produced by an enzyme cascade system (four enzymes and a specific substrate) whenever a nucleotide is complementary to the template.	Allow rapid real-time determination of 20–30 bps of target DNA.	Template preparation is required. Limited multiplexing capability.	Ronaghi et al. (1998)

Continued

Table 13.1 Summary of major SNP genotyping assays along with principle and detection method for each allelic discrimination reaction.—cont'd

S.no	SNP typing assay methods	Principle	Detection method	Methodology	Advantages	Disadvantages	References
3	**Allele-specific oligonucleotide (ASO) ligation**	Based on the capability of DNA ligase in joining two oligonucleotides when they hybridize adjacent to one another on a DNA template. Two types of reactions exist:					Landergren et al. (1988), Barany (1991)
		1. Ligase chain reaction. Three probes—one common probe annealed to the target DNA; the other two ASO probes competed to anneal to the DNA target, resulting in a 'nick' at the allelic site. DNA ligase will ligate	1.1 Fluorescence	Biotin on the common probe and reporter group on the allele-specific probe can be used to detect the ligated product.			Landegren et al. (1988)

		the annealed allelic probe to the common probe. Ligation products are amplified using PCR.					
		2. Sequence-coded separation reaction	2.1 Electrophoresis	Mobility modifiers and fluorescent dye-labelled allelic-specific probes enables multiplexed OLA, where ligation products are detected using a fluorescence DNA sequences.		All the reagents are added simultaneously reducing the manual work and allowing easy automation.	Grossman et al. (1994)
			2.2 FRET (With the ligation, an increase in FRET is detected)	Three dye-labelled ligation oligonucleotides, where the donor dye labels the 5′ end of the common probe while the acceptor dye labels the 3′ end of the ASO probes.			Chen et al. (1998)

Continued

Table 13.1 Summary of major SNP genotyping assays along with principle and detection method for each allelic discrimination reaction.—cont'd

S.no	SNP typing assay methods	Principle	Detection method	Methodology	Advantages	Disadvantages	References
4	**Invasive cleavage/ invader assay**	Based on recognition and cleavage by a flap endonuclease when two oligonucleotides invader nucleotide and the probe anneals to the target DNA such that the probe overlaps the 3′ end of the invader nucleotide, forming a 3D structure that is cleaved by the flap endonuclease, releasing the 5′ arm probe.	4.1 FRET	A reporter dye at 5′ arm and quencher complementary to the probe. With the cleavage, the fluorophore is removed, and fluorescence is enhanced.	No requirement for previous PCR amplification.	Target DNA is needed in large amounts.	Lyamichev et al. (1999), Kaiser et al. (1999), Mein et al. (2000)
			4.2 Fluorescence polarization (FP)	• Molecular weight of the fluorophore-labelled probe decreases with the decrease in FP on cleaving.	Increase in sensitivity with PCR before the invader reaction.		Hsu et al. (2001)
			4.3 Mass spectrometry	Different numbers of nucleotides are used to design probes.			Griffin et al. (1999)

terms of output and overall expense, thus being overtaken by next-generation sequencing (NGS) technology alternatively massive parallel sequencing (MPS) (Kircher & Kelso, 2010).

NGS is defined as a nonSanger-based technology that allows the parallel sequencing of millions or billions of DNA molecules, resulting in substantially high throughput with the minimal requirement for cloning of fragments commonly employed in Sanger sequencing (Sobiah et al., 2018). This massively parallel sequencing is obtained by the miniaturization of each sequencing reaction, reducing the size of the instrument along with the reagent cost per reaction (Qin, 2019). It includes second- and third-generation sequencing platforms in which miniaturization has first allowed the analysis of a large number of samples simultaneously and then the determination of the sequence of single DNA molecules (Sobiah et al., 2018). A significant trait of all NGS platforms is the limited read length in the range of hundreds of base pairs, which requires material to be fragmented prior to analysis if the DNA to be sequenced is lengthier than the suitable read length. At the end of sequencing, the reads are reassembled *in silico* to achieve the whole sequence of the target molecule (Qin, 2019). However, second-generation sequencing technology holds the drawback of having relatively small read lengths, leading to complexities in splicing, assembly, and annotation (van Dijk et al., 2014), whereas third-generation sequencing technology holds the advantage of detecting single molecules along with real-time sequencing.

Current next-generation sequencing platforms

The use of NGS has offered a concurrent examination of forensically admissible genetic markers, including SNPs, mutations, STRs, and transcripts through massively parallel sequencing (Sobiah et al., 2018). It has been implemented mainly in the identification of humans along with phenotypic trait determination. Different sequencing platforms to present are Roche (454) sequencing, Illumina sequencing, Applied Biosystems (SOLiD) Sequencing, Ion Torrent (Thermo Scientific), Helicos BioSciences Corporation (Heliscope) sequencing, Pacific Biosciences SMART Sequencing, Oxford Nanopore, Qiagen, and BGI (Goodwin et al., 2016).

Two dedicated systems among the MSP platforms that are used in various forensic laboratories are MiSeqFGx (Illumina) and the Ion GeneStudio S5 system (Thermofisher Scientific). The former is one of the most widely used MPS for forensic casework with the ForenSeq DNA signature panel. The kit contains more than 200 forensically relevant genetic markers, consisting of 27 autosomal STRs, 24 Y-STRs, 7 X-STRs, 94 iSNPs, 56 aSNPs, 24 pSNPs, and amelogenin (Jäger et al., 2017). Ion Torrent's GeneStudio S5 system is also broadly acknowledged for forensic casework using Precision ID Identity Panel Kit, which contains 90 autosomal SNPs and 34 Y upper-clade SNPs in a single, efficient workflow (Dash et al., 2022).

All these platforms are based on the principle of SBS technology and allow forensic scientists all over the world to utilize the maximum potential of NGS. With attributes such as read length, number of reads, run time, advantages, and disadvantages, a platform comparison between the characteristics of current NGS platforms is described in Table 13.2.

The major steps for sequencing are similar for all these platforms: Construction and amplification of DNA templates; distribution of templates on a solid support; sequencing and imaging; base calling, quality control; and analysis of data. In regard to operation, two major approaches are taken by NGS: de novo sequencing and resequencing. In the first, the genome sequencing of an organism is done for the first time, while in the second, the genome of a species is sequenced and a reference sequence is readily available. On this basis, a strategy of sequencing is selected, and data analysis is done. The approach of resequencing is followed in Human forensics and population genetics studies, while in microbial forensics, the two approaches of de novo sequencing and resequencing of microbial genomes are followed.

NGS versus traditional capillary electrophoresis-based assays

Although CE-based assays involving SNPs and STR genotyping have become the gold standard for DNA sequencing in forensic samples over the past 20 years, they hold several significant disadvantages.

1. Foremost drawback of CE-based assays is their inability to multiplex. It becomes complex to simultaneously detect more STR markers, resulting in overlap and size constraints for fluorescently labeled amplicons (Oostdik et al., 2014; Wang et al., 2015). Due to this incapability of multiplexing, samples are processed individually. This makes the sequencing tedious and error-prone.
2. Another limitation is the low-resolution genotyping of current markers. There is uncertainty in polymorphic allelic capture having variants of substitution as traditional CE-based analysis is based on length-based typing; henceforth, identification of sequence differences present in alleles of the same length is not possible. Therefore, mixed DNA sample analysis could not be carried out, which reduces the resolution of the technology. Such compromised samples yield low detection rates, leading to a loss of useful genomic information. Moreover, traditional CE-based assays cannot determine mutations in complicated paternity cases.

Advantages of NGS for STR and SNP genotyping

All the novel sequencing platforms have been directed toward four main refinements over the traditional technologies. Firstly, these platforms rely on the construction of NGS libraries in a cell-devoid system, avoiding bacterial cloning of DNA fragments. Secondly, thousands to millions of sequencing reactions can be parallelized with NGS. Also, analysis of multiple polymorphisms can be done efficiently, which was earlier a limitation

Table 13.2 Comparison of characteristics of current NGS platforms.

Manufacturer	Instrument	Chemistry	Read length (nucleotides)	No. of reads	Output (Gb)	Runtime	Advantages	Disadvantages	Applications
Illumina established in 1998, San Diego, CA	MiSeq FGX	SBS	2 × 300 bp	25 million	15 Gb	4–55 h	Short run time and ease of use.	Expensive per base	Long-read applications
	NextSeq 550 series		2 × 150 bp	400 million	120 Gb	12–30 h			Exome and transcriptome sequencing
	NextSeq 1000 and 2000 series		2 × 150 bp	1.1 billion	330 Gb	11–48 h	High yield		Mi RNA and sRNA analysis
Ion Torrent launched in 2010 by life technologies corp (now part of thermo Fisher scientific)	Ion Gene studio S5 system	SBS	200 –600 bp	2–80 million	15 Gb	4.5–21.5 h	Efficient, scalable, and low cost targeted sequencing	Yields fewer reads at a higher cost per Mb	WGS (Whole genome Sequencing), TGS (Target gene sequencing)
	Ion Gene studio S5 Plus system		200 –600 bp	2–130 million	30 Gb	3–20 h			Prepare library for AmpliSeq, reproducible template and to load chip.
	Ion PGM 316 chip		200 –400 bp	2–3 million	300 Mb –1 Gb	3–4.9	Short run time and low reagent cost. Economical for the lowest sample input		Regulated lab environment, in vitro diagnostics, sRNA and de novo sequencing, SNP verification.
	Ion Torrent GeneXus		200 bp	12–15 million	15 Gb	2.5 h			WGS, termed as in-house NGS system
Roche 454 1st commercial systems was launched in 2004.	Roche 454 GS FLX	SBS	700 bp	1 × 106	0.7 Gb	23 h	Long reads, short run time	Homopolymer errors, expensive	Used in de novo sequencing projects. Meta genomics
	Roche 454 GS FLX titanium system launched in 2004		750 bp	1 million	0.7 Gb	24 h			

Continued

Table 13.2 Comparison of characteristics of current NGS platforms.—cont'd

Manufacturer	Instrument	Chemistry	Read length (nucleotides)	No. of reads	Output (Gb)	Runtime	Advantages	Disadvantages	Applications
GenapSys founded in 2010 is a company from stanford Genome Technology center	GenapSys (16 chips)	SBS	150 bp	13 million	2 Gb	24 h	Low priced, easy to use	Low performance	Identifying pathogens, RNA, sWGS, targeted mRNA, SCP and gene editing.
Applied biosystems SOLiD 1st SOLiD platform was released in 2007	SOLiD5500xl	Sequencing by ligation	75 bp	1.5 × 109	180 Gb	14 days	Lower cost and inherent error correction	More run time, short reads	Whole genome sequencing
Pacific biosciences was launched in 2010.	PacBio (HiFi reads)	Real-time sequencing	25 kb	1 × 105	66.5 Gb	30 h	Long reads, short run time	High error rate, low yield	Characterizing bacterial genomes and in human disease related studies.
Nanopore fourth generation sequencing	Nanopore (PromethION)	Real-time sequencing	4 Mb		14 Tb	72 h	Spatial distribution of DNA reads, large sample size (1—48 flow cells)	Not evaluated	Identifies biomolecules while they are passed through nanoscale holes, personalized medicine

in CE-based genotyping (Churchill et al., 2016). Thirdly, no electrophoresis is required for the detection of sequencing output. Lastly, at an unprecedented speed, NGS generates an enormous number of reads that enable the sequencing of entire genomes. Besides this, as already mentioned NGS typing is based on SBS chemistry (Bentley et al., 2008), which results in benefits like error removal, low incorporation bias, and missed base calls in the homopolymers and repetitive regions (Bragg et al., 2013). A prominent advantage with NGS is the high level of sensitivity with digital data in spite of the analog metrics like shape, size, height, and color of the peak resulting from traditional CE-based systems. Quantitative data is developed by NGS in the form of classifiable read counts. Moreover, sensitivity can also be increased with an increase in coverage level and in-depth overall sequencing, which aid in achieving the small contribution of donors in a complex mixture sample. This outcome cannot be achieved with a CE-based assay.

Although the most accepted STR genotyping technique provides satisfactory discrimination power for a wide range of forensic applications, NGS-generated STRs hold many potential advantages in the forensic analysis of challenged samples. Some of the advantages of NGS for the analysis of STR are high throughput, concurrent detection of a large number of STR loci on both autosomes and allosomes, and the capability of differentiating alleles with similar digital read counts. Also, NGS technology facilitates the analysis of mixed DNA samples and complex paternity cases, eventually upgrading the workings of legal cases at a lower cost. STR allele calls generated using NGS are wholly in accordance with the current database format, establishing an ideal relationship between NGS and CE-based data. NGS has also become a tool for differentiating between mutations and exclusions, with more STR loci being detected simultaneously. Meanwhile, NGS technology identifies the sequence of the STR loci so that the variation in the repeat within the STR loci can be detected, which may conclude the origin of mutation, thereby increasing the power of discrimination.

SNP detection possesses the explicit advantage of NGS application in the realm of forensic genomics. It offers the capability of differentiating alleles similar in length but differing in sequence. Over the years, forensic DNA phenotyping (FDP) employing SNP genetic marker has been used extensively for criminal investigation when STRs are insufficient. The markers work along with bio-geographical ancestry (BGA) and morphologically visible characters (Kayser, 2015). Compared to STR markers, they possess lower mutation rates—about 100,000 times lower than STRs and are thus more stable. SNPs are unlikely to change over time, stating that they remain the same their whole lives (Biesecker et al., 2005). This makes them more favorable to be used in kinship analysis, pedigree or lineage reconstruction cases, and mass disaster victim identification. SNPs can work effectively with compromised biological samples with low-quantity DNA templates and/or degraded samples. In such compromised samples, STR analysis may not yield a better result due to the requirement of a larger intact fragment of DNA. With a small fragmented DNA length of about 60–80 bps, SNP analysis

can be performed (Budowle et al., 2004; Divne and Allen, 2005). The study of phenotypic characteristics such as hair and eye color has also improved with SNP genotyping through NGS. Since most SNPs are biallelic, it decreases the informativeness per locus when compared to STRs. However, this limitation can be easily compensated with a much larger set of SNPs using the NGS technique to accomplish the same power of discrimination afforded with the core STR loci. A set of about 50–100 SNPs is equivalently discriminating as a set of 10 STRs (Chakraborty et al., 1999). Table 13.3.

Forensic application of SNP markers using NGS technology

Analysis of highly degraded and mixed samples

In routine forensic DNA analysis, sample quantity and quality are major concerns. In conventional techniques using STR markers, analysts encounter problems majorly from highly degraded and contaminated DNA recovered from challenging samples, i.e., samples from highly decomposed human corpses, cremated dead bodies, ancient skeletonized bodies, charred bones, teeth, and hair shafts exposed to harsh environmental conditions (Carrasco et al., 2020). However, the emergence of advanced technology in

Table 13.3 Summary of comparison of characteristics of STRs and SNPs.

Characteristics	STRs	SNPs
Mutation rate	10^{-9} nucleotide per generation	10^{-6}–10^{-2} nucleotide per generation
Analysis of degraded samples	Less suitable	More suitable
Power of discrimination	High	Low; but same level of discrimination can be achieved by multiplexing
Mixture analysis	Poor resolving power	High resolving power due to large number of markers and SNP microhaplotypes.
Rare polymorphism	Cannot be studied	Can be reliably detected by analyzing every base of the genome.
Mitochondrial DNA analysis	Detects polymorphism only within a hypervariable region.	Potential to analyze whole mitochondrial sequence and heteroplasmy is detected (Holland et al, 2011).
Ethnicity estimation, physical traits, etc	Not reported	Reported with an accuracy of 90%.

the form of NGS opened new possibilities for human identification from challenging and compromised samples. In the case of degraded samples, SNP is a better potential marker for human identification than other available technologies because it generates short read lengths and amplicons (<100 bp) (Butler, 2011, pp. 347−362). Therefore, it lowers the error rate, which is a major advantage while dealing with degraded samples. To deal with degraded samples, there is an urgent need to employ new approaches in forensic casework analysis. In the past few years, various NGS methods have manifested efficacy with compromised samples (Cho et al., 2020; Turchi et al., 2020) and mixed samples with unequal ratios of DNA donors (Chen et al., 2020; Hwa et al., 2020; Momota et al., 2021; Petrovick et al., 2020; Ricke et al., 2021). The human identification SNP multiplex assay, i.e., ID 52-plex, was developed using 52 unlinked bi-allelic SNPs in European, Asian, and African populations and carried out DNA genotyping. The study reported that 52 SNP loci were efficiently amplified with a mean match probability of 5×10^{19} from 700 degraded samples of the studied population, while partial STR profiles were generated from the CE method (Sanchez et al., 2006). Therefore, SNP proved to be a more efficient marker than STR in the case of analysis of degraded DNA samples (Sanchez et al., 2006). Cho et al. (2020) tested 70-year-old postmortem samples of human skeletal remains using the MPS-based STR system, which is recommended best for degraded samples, and reported the higher success rate of the MPS-based STR system compared to the CE-based STR typing method. Another study showed the implication of NGS technology by analyzing degraded vs. non-degraded and mixed samples of two close relatives in varied ratios (minor to major contributors), i.e., 1:9, 1:19, 1:29, 1:39, 1:79, and 1:99 (Hwa et al., 2020). Genotyping results of MPS-based 59-identity STRs and 94-identity SNPs showed that 82.4% and 54.5% of minor contributors could be precisely recognized from both intact and degraded DNA mixtures, respectively (Hwa et al., 2020). Additionally, micro-haplotype assays have also been developed to emphasize unequal contributors to the DNA mixture (Bennett. et al., 2019; Chen et al., 2019).

Of 273,566 multi-allelic SNPs (from the 1000 Genomes Phase III) located across the human chromosomes (Phillips et al., 2020a, 2020b), approximately 99% are triallelic and ~ 1% are tetra-allelic SNPs. Compared to the biallelic SNPs, these SNPs have high discrimination power and are helpful in analyzing highly degraded and mixed DNA patterns (Phillips et al. 2020b). A large-scale MPS-based tetra-allelic SNP panel has been designed by selecting 160 SNPs with high heterozygosity and 24 SNPs with high discriminating power. This panel enables the detection of more than two alleles in at least one marker in 99% of mixed sample cases belonging to African and European donors, while it is reduced to 92.6% in the case of Asian donors (Phillips et al., 2015). A SNaPshot approach-based panel, the multiple-allele SNP test for forensics, containing 27 triallelic and 2 tetraallelic SNPs was designed for forensically degraded samples and DNA mixture deconvolution. It was presented with high discrimination power among all five main population groups, with a random match probability value ranging from 10^{-15} to 10^{-20}

(Phillips et al. 2020b). This small panel's information could be used for the development of large-scale MPS-based multi-allelic SNP panels that are thought to be more potent for degraded and DNA mixture deconvolution.

With the availability of refined genotyping techniques for the isolation of DNA from degraded skeletal material, we are still facing a lack of sequencing success due to the low DNA quantity and quality of compromised samples. A recent study recommended the implementation of optimization strategies in sequencing workflow while dealing with highly degraded samples (with concentrations $\geq$31.2 pg), such as increasing the volume of the library pool, performing additional PCR purification, and reducing the adapter volume (Senst et al., 2022). Prior to massively parallel sequencing, target enrichment approaches using biotinylated probes have also shown remarkable results in the case of degraded and ancient DNA (Bardan et al., 2023). The power of advanced genetic technology can be evidenced by the recent discovery of Scientist Svante Pääbo, awarded the "2022 Nobel Prize in Physiology or Medicine," who decoded the mtDNA from a 40,000-year-old archaic bone sample of Neanderthal, a species of humans that went extinct (Vernot et al., 2021).

Low-copy DNA detection with NGS approach

Presently, NGS platforms are forensically validated with multiple panels having different sensitivities for input DNA varying from 62.5 pg to 1 ng for autosomes, X and Y chromosome SNPs, and STR (Carratto et al., 2022). The first commercially available MPS STR panel for the MiSeq FGx platform is ForenSeq DNA Signature (Illumina, CA), which includes 172 SNPs, 27 autosomal STRs, 7-X STRs, and 24 Y-STRs. Jäger et al. (2017) validated the kit and reported 89% of complete STR profiles of 26 autosomal loci (out of a total of 27 loci) from all dilutions of DNA ranging from 1 ng to 62.5 pg. The *D22S1045* marker was excluded from this study due to its poor sensitivity, whereas X and Y STRs showed 100% generation of profiles with 1 ng input DNA. Besides, no significant decrease in sensitivity was found up to 62.5 pg of input DNA. Input DNA quantities below 8 pg depicted more than 50% loss of SNP calls. The sensitivity of the Identity Panel (which includes 124 SNPs) for human identification is as low as 100 pg of DNA, while the Ancestry Panel, which encompasses 165 biogeographic ancestry SNPs, requires 125 pg of DNA. The GlobalFiler NGS STR Panel includes 35 STR (autosomal, X, and Y chromosomes), which is an important panel for mixed samples in varied ratios and requires a minimum of 125 pg of input DNA. A panel for combined forensic phenotyping and ancestry determination, the VISAGE-Basic Tool Research Panel (which includes 152 SNPs and 1 Indel, of which 41 SNPs belong to the HIrisPlex-S system) generates a profile with 1 ng of template DNA. Its validation study reported the generation of a full profile at a dilution of 100 pg of DNA (Xavier et al., 2020). An all-in-one SNP panel, FORensic Capture Enrichment (FORCE), comprising

5422 forensically relevant SNPs (identity, ancestry, phenotype, and lineage SNPs), was used to generate a profile from the 200-year-old bone samples. 44.4% of the SNPs present in the panel could be generated successfully from such an old bone sample (Tillmar et al., 2021).

Individual identification

SNP markers with high power to differentiate among individuals are known as identity-testing SNPs. It represents the same function as conventional Combined DNA Index System (CODIS) STR markers but with different discrimination power, sensitivity, and specificity. SNPs with the highest heterozygosity (50% heterozygosity) and low Fst values are considered the ideal identity markers. Kidd et al. (2006) screened 90,000 candidate SNPs in European Americans, African Americans, and Chinese/Japanese and selected 2723 markers with average heterozygosity ≥ 0.45 and Fst < 0.01 for the three populations. Among 2723 SNPs, another 35 were chosen after testing in a broad population set, and finally, 19 identity SNPs were selected with a high value of heterozygosity and a low Fst in all 40 populations. SNP markers are identified as an efficient profiling tool for individual identification with the application of advanced throughput techniques. Several NGS-based kits are commercially available for CODIS STR loci used for forensic testing. Currently, major NGS platforms available in the market are Illumina's MiSeq (San Diego, CA) and the Ion S5 PGM System (Thermo Fisher) which are forensically validated with various MPS kits for CODIS STRs as well as NGS SNP multiplex panels for individual identification, such as the ForenSeq DNA signature panel (Illumina), the MPS panel for SNP for ID 52-plex (Valle-Silva et al., 2019) and the 92-Plex IISNPs, SNP panel for 140-locus (QIAGEN) (Avent et al., 2019), the 21-plex STR identity panel (Silvery et al., 2020), and the SifaMPS panel comprising of 87 STRs and 294 SNPs (Tao et al., 2021). NGS panels with Precision ID identities for 124 SNPs, a 448-plex SNP panel (Zhao et al., 2021), and a multiplex panel of 1270 triallelic SNPs (Phillips et al., 2009) are also in use. Multiple studies have demonstrated that autosomal and Y-STR analysis using MPS panel-based genotyping-by-sequencing (Syndercombe-Court, 2021; de Knijff, 2022) is a more robust and sensitive technique than CE-based genotyping-by-length approaches. NGS STR panel Precision ID identity (Thermo Scientific, US) has been developed for human individualization using 124 identity SNPs (comprising 90 autosomal and 34 Y-chromosomes). A validation study has assessed these identity SNPs in the Central Indian population, on the Ion GeneStudio S5 instrument and reported that the allelic frequencies of SNPs ranged from 0.001 to 0.377 in autosomal SNPs. A SNP, *rs9951171*, was reported to be the most potent marker in the Central Indian population with the highest power of discrimination and the lowest cumulative match probability value of 4.76698×10^{-37} (Dash et al., 2022b). The high-throughput NGS technique has major importance in solving the discrimination of closely

related individuals by sequencing and scanning the mutations in flanking regions of STR loci, which elucidate the newer alleles that ultimately increase the discriminating power (Rockenbauer et al., 2014; Devesse et al., 2018). In the Indian scenario, Dash et al. (2021) conducted the first study conferring sequence-based allele determination in the Central Indian population utilizing the NGS STR panel Precision ID GlobalFiler v2. A subsequent increase in allele number was observed in 31 autosomal markers due to mutations found in the flanking region of STR loci and also revealed a most potent substitution or mutation *rs4847015* at trinucleotide STR marker *D1S1656* in 11.5% of the Central Indian population (Dash et al., 2022). Allele gain at intervening regions as well as flanking regions at multiple STR loci increases the statistical or discrimination power, leading to a higher evidentiary value of NGS-based SNP analysis compared to CE-based STR analysis (Dash et al., 2021). Additionally, multiallelic SNPs (triallelic or tetraallelic) have a high genotypic combination, thereby increasing the level of polymorphism and subsequently the power of discrimination value compared to the biallelic markers (Cao et al., 2015; Cornelius et al., 2019; Gao et al., 2018; Phillips et al., 2015; Westen et al., 2009). SNPs *rs4540055* and *rs5030240* are reported to be the most potent triallelic forensic markers for individual identification (Cao et al., 2015; Westen et al., 2009). In combination with SNPs, Insertion deletions (InDels), which are the second most abundant biallelic polymorphisms (approx. 3.5 million InDels) in the human genome (The 1000 Genomes Project Consortium et al., 2015). Considering the lower mutation rates ($\sim 2.9 \times 10^{-9}$) (Nachman and Crowell, 2000) and shorter amplicon sizes resulting in a low rate of ambiguous sequencing results such as stutter generation, therefore InDels are also deciphered as another useful marker for forensic cases than STRs (Al-Snan et al., 2021; Zaumsegel et al., 2013).

Human ancestry and migration history

Although identity SNPs and STR loci are powerful markers for individualization, they are inadequate for elucidating the biogeographical evolutionary ancestry and migration history of an individual due to the high chance of allele-sharing of these markers among populations (Zheng et al., 2018). Approximately 80%–90% of the genetic variations are shared among all individuals, while only 10%–20% of the total variation is population-specific, which can be referred to as "private variation" in a population (Deka et al., 1995; Nei, 1987). Allelic frequencies of these ancestry markers significantly vary among different continental and subcontinental populations (Kosoy et al., 2009). A commercial kit, DNA Witness BGA was developed to elucidate the ancestry of four continental populations, viz., American, European, East Asian, and sub-Saharan African, by calculating the proportion of different ancestry components in admixed individuals (Frudakis, 2008; Shriver et al., 1997). Furthermore, two more array-based genotyping panels were developed: The first panel consisted of 320 Eurasian-specific ancestry markers

that determine admixture among Southeastern European, Northern European, Middle Eastern, and South Asian populations (Frudakis, 2008), and the other panel consisted of 1476 European ancestry markers that have the power to elucidate European subancestry such as Jewish, Italian, German, Irish, Finnish, Greek, Spanish, and Russian, etc. (Bauchet et al., 2007; Frudakis, 2008). The VISible Attributes through GEnomics (VISAGE) Basic Tool (BT) is the first forensic MPS single assay test comprising 120 markers for predicting both external visible characteristics (EVCs) as well as continental-based ancestry. In 2020, a consortium VISAGE expanded the use of MPS-based techniques for predicting the biogeographic ancestry, facial phenotype, and age of an unknown accused sample (Palencia-Madrid et al., 2020). Bardan et al. (2023) developed a customized hybridization capture panel that detected 75% of correct predictions of the European biogeographic ancestry of individuals and 83% and 92% of correct predictions of hair and eye color, respectively (Ruiz et al., 2013). In recent years, a number of BGA panels compassed in predicting the genetic ancestry of the population at the continental level (Gettings et al., 2014; Lao et al., 2006; Pereira et al., 2019; Phillips et al., 2013; Qu et al., 2019; Santos et al., 2016; Xavier et al., 2020; Zhao et al., 2019). However, we are faced with compromised accuracy in determination of biogeographic ancestry when studying strictly geographically confined populations, which may be due to candidate SNP selection and their limited number (Mehmood et al., 2012). Recently, a study reported the enhanced capacity of concluding ancestry at the continental and subcontinental levels by using four forensic SNP panels with the implementation of an algorithm for partial least squares discriminant analysis (PLS-DA). In line with this study, Pilli et al. (2023) developed a new BGA panel comprising >3000 SNPs and evaluated the accuracy of the panel using the PLS-DA method. By using different selection techniques and genetic algorithms, we reduced the number of markers by selecting the 800 most informative SNPs to infer the ancestry of populations at continental and subcontinental scales.

With the fact that a phenotype is determined by both genotype and environment when an ancestral population migrates to an extreme environmental or geographical location, it reshapes a population's ancestry under the influence of genetic adaptation to a microenvironment due to the process of natural selection (Harris & Pritchard, 2017). An Indian study of whole genome sequencing (WGS) of a highland native from Ladakh, an individual from submountaneous Kumaun region, and an individual from Telangana (Andhra Pradesh) revealed high altitude-relevant genetic variants of *EGLN1 [rs12097901 (Cys127Ser)* and *rs186996510 (Asp4Glu)]* in Ladakh natives (Malhotra et al., 2018). This corroborates with many studies of *EGLN1* being positively selected for high altitude adaptation (Lorenzo et al., 2014; Xiang et al., 2013). In a forensic context, these adaptation loci could also be useful in establishing a person's biogeographic history, thus helping to concretize the possible matches from a list of missing individuals.

Facial phenotyping

Recent developments in genotyping or sequencing techniques have proven SNPs to be a very informative tool in determining a person's face EVCs or phenotypic traits such as skin color, hair color, eye color, and face morphology and hair morphology. EVCs are polygenic in nature, which is determined by the complex interplay of an individual's genotypes as well as environmental factors (including altitude, latitude, diet, exercise, and stress exposure such as UV exposure, heat, cold, oxygen level, etc.). However, by applying advanced high-throughput techniques, additive genotype information of various phenotypic markers helps a forensic scientist predict the physical appearance of unknown perpetrators based on a DNA sample found at the crime scene (Kayser, 2015; Seo et al., 2013), which provides a lead to investigating agencies and narrows down the list of suspects. This inferential technique is known as FDP. The FDP technique would be especially useful in forensic cases where there is no reference sample available for comparison with an unknown DNA sample.

Human pigmentation phenotypes such as hair color, eye color, and skin color have been extensively studied. Besides these traits, other polygenic phenotypic traits such as face morphology, freckles, age, height, and myopia have been rarely studied. The variation in human skin, eye, and hair color is due to the amount of production and two types of melanin pigment involved in the pigmentation pathway. The melanin pigment type that is responsible for a darker (brown/black) phenotype is called eumelanin, and another pigment responsible for a lighter (reddish) color is called pheomelanin. Some candidate genes, viz., *OCA2, HERC2, SLC24A5,* and SLC45A2, encoding membrane ion transport proteins have shown a significant role in the amount of melanin production by affecting the tyrosinase activity encoded by *TYR*. Three SNPs located on the *OCA2* gene, viz., *rs6497268, rs7495174,* and *rs11855019*, have been shown to have a strong association with blue versus nonblue eye color (Duffy et al., 2007). Moreover, another association study reported the strongest significant association of green iris color with OCA2 marker rs1800407 and two important eye color variants (*rs74653330* and *rs121918166*) (Branicki et al., 2008; Andersen et al., 2011). Except for *OCA2*, a genome-wide association study (GWAS) has observed up to 15 SNP markers in HERC-2 gene. Among them, *rs916977* (where the *T* ancestral allele is associated with brown iris color and the *C* allele is associated with blue iris color) and *rs1667394* emerged as important markers in the determination of eye-color variation in European descent (Kayser et al., 2008). A number of phenotype-genotype association studies helped in the development of efficient and reliable multiplex genotyping panels such as IrisPlex for prediction of eye color (blue/brown) consisting of six potent markers in different loci, viz., *OCA2 (rs1800407), SLC24A4 (rs12896399), SLC45A2 (rs16891982), HERC2 (rs12913832), IRF4 (rs12203592), and TYR (rs1393350)* based on the European (Dutch) population genotype and phenotype data (Ruiz et al., 2013; Yun et al., 2014).

These phenotype-related markers were further scanned in other European countries (including Norway, Estonia, the UK, France, Spain, Italy, and Greece), manifesting high predictive power (area under curve = 0.96) for two eye colors (blue and brown) (Walsh et al., 2011). It was observed that for intermediate eye colors, IrisPlex could not give as accurate a prediction as with blue and brown eyes (Liu et al., 2009; Walsh et al., 2011, 2012). As per the available studies, along with *OCA2 rs180040, HERC2 rs12913832* was found to be the most potent marker, with highest prediction values ranging from 68.8% to 74.8% for eye color variation (Andersen et al., 2011; Valenzuela et al., 2010). However, the selection of the ideal SNP markers for the phenotypic prediction model appears to be highly population-specific. In the Polish population, *rs1408799* was significantly associated with eye color, *rs1800416* in the Brazilian population, and *rs916977* in the Czech population (Andrade et al., 2017; Zidkova et al., 2013). Genetic variations associated with skin pigmentation were observed among individuals belonging to biogeographically distinct continental countries of the world (Asian, African, Caucasian, and Australian) or also among individuals belonging to the same geographic region depicting a continuous variation in skin color shade.

Genetic polymorphisms associated with hair pigmentation variation are *SLC45A2 rs16891982, SLC452 rs26722, and ASIP rs369378152* (van Daal, 2008). Among them, significant allele frequency variation was observed for *SLC45A2* (rs16891982) in dark-haired nonCaucasian continents (viz., Asians, Americans, and Africans) and light-pigmented Caucasian populations (van Daal, 2008). A cSNP *rs3827760* in the gene ectodysplasin A receptor was identified as the dominant marker for straight hair in East Asians (Riddell et al., 2020). With time and increased available genotype phenotype data, another HIrisPlex model was developed for combined prediction of eye and hair color, consisting of 22 most predictive SNPs (4 markers of IrisPlex+18 new markers). The HIrisPlex tool has shown BGA prediction accuracies of 69.5% for blond hair color, 78.5% for brown, 80% for red, and 87.5% for black independently in the European population. However, when it was tested on a nonEuropean population, it demonstrated satisfactory eye/hair color prediction but with minor accuracy (Balanovsky et al., 2019). Moreover, a study identified *HERC2 rs1129038* as a potent marker for the prediction of blond hair, *HERC2 rs12913832* for black hair, and *HERC2 rs12931267* and *TPCN2 rs35264875* as ideal markers for brown-colored hair. Later on, the third-most important combined model, HIrisPlex-S was created for concurrent prediction of eye, hair, and skin color, encompassing 36 markers (of which 19 markers from HIrisPlex 17 novel markers). Compared to other models, the HIrisPlex-S tool proved to have high prediction accuracy and discriminating power among different tones of skin on a global scale, i.e., 72% accuracy for pale skin, 74% for very pale, 73% for intermediate, 87% for dark, and 97% for dark black (Chaitanya et al., 2018). The compiled phenotype and genotype data from IrisPlex, HIrisPlex, and HIrisPlex-S systems are publicly available in an interactive tool (https://hirisplex.erasmusmc.nl/). Through this tool, prediction

models can generate individual probabilities for available categories of five skin colors, three eye colors, and four hair colors from genotype data for 41 markers. Alternatively, the physically visible appearance of an individual may also infer his BGA.

In the context of facial phenotyping, a vast literature is available on the application of GWAS rather than NGS techniques on a genomic scale. Over the last decade, multiple GWAS studies have revealed various genetic markers associated with facial shape phenotypes (Xiong et al., 2019) and syndromes associated with abnormal facial features such as Down syndrome, Prader-Willi syndrome, and cleft lip (Boehringer et al., 2011; Brinkley et al., 2016). The application of high-throughput NGS-based genotyping techniques has considerably accelerated the process of causal gene discovery in the case of polygenic complex traits with effective prediction accuracy. Another robust technique whole exome sequencing (WES) allows sequencing of only the protein-coding region (known as the exome), which forms 2% of the whole genome (Ng et al., 2009). Approximately 80% of mutations in the exome region cause Mendelian diseases. In the field of medical genetics, the WES approach has become a robust technique for the prognosis of rare genetic disorders that have a prevalence below 1/2000 (Ng et al., 2010). Therefore, the application of WES or WGS can prove an important technique for the diagnosis of craniofacial Mendelian traits or genetic syndromes related to facial feature abnormalities through the unraveling of rare pathogenic variants (Al-Nbaheen, 2018; Yang et al., 2013). Another recent phenotyping application, Face2Gene (http://www.face2gene.com) is available for clinical and research purposes and is basically useful for the evaluation of genetic syndromes related to facial dysmorphology by using patient information such as photo, gender, weight, height, etc. Limited studies are available in the context of phenotypic and genotypic association studies in the Indian population. Extensive research is needed for the discovery of robust candidate markers, new MPS panels/assays, and their validation studies in different populations for these phenotypic markers.

Paternal lineage and kinship analysis

Paternal lineage and kinship-associated SNPs have been extensively studied in the X chromosome and uniparental Y chromosome, as well as in mtDNA. Lack of recombination and a lower mutation rate (10^{-8}) than STR markers (10^{-3}) are the most important features of these lineage markers (Amorim and Pereira, 2005). In forensics, lineage SNPs have proven advantageous in the kinship analysis of an individual in cases of mass disaster identification or missing person identification, mostly in the conditions of the unavailability of close first-degree relatives as reference samples.

Numerous studies are available on the validation and assessment of various Y-STR typing systems, such as a 24 plex (Fan et al., 2021), a 41 plex (Liu et al., 2020), a multiplex with 12 multicopy Y-STR loci (Shang et al., 2021), and a 26 plex for rapidly mutating Y-STRs (Albastaki et al., 2022). Such studies have proven the importance of lineage SNPs

in sexual assault and paternity lineage differentiation cases. Nevertheless, limitations of these CE-based commercial Y-STR multiplex kits are being faced in the identification of closely related males belonging to the same paternal lineage and mixed samples of incest cases due to the unavailability of a sufficient number of loci in any single amplification kit. In such situations of inconsistent results from CE-based STR typing, genetic markers such as Ins/Del (González-Herrera et al., 2020), haplogroup analysis, and SNP markers (Børsting & Morling, 2011) have been identified as alternative markers in complex paternity cases. In this regard, *Penta E* was found to be the most potent STR marker for paternity determination in the Indian population, with heterozygosity (0.906) and the lowest value of match probability (0.022) (Dash et al., 2021).

Multiple studies are available on NGS technology-based multiplex panels used for paternity and kinship analysis. In many father-child duo cases, the 49 SNP genotyping information of the 52-plex test helped in assessing the paternity, while the STR profile was found to be insignificant in proving the paternity of the alleged father. Moreover, in an incest case, the grandfather was convicted as the real biological father, and the putative father was proved innocent by the SNP analysis (Børsting & Morling, 2011). Hence, in addition to STR studies, these cases showed the impact of SNP markers in assessing complex paternity and kinship cases. The World War II case of family relationship was solved using the all-in-one SNP panel FORCE, which encompasses 5422 forensically relevant markers (X, Y SNP, ancestry SNP, and phenotype markers) by predicting the distant relationships (up to the fifth degree) with strong statistical results (Tillmar et al., 2021).

Y-SNP haplogroup studies utilizing high-resolution NGS approaches to enhance our understanding of male individuals in forensic genetics and evolutionary studies. Dash et al. (2022) NGS study revealed 11 unique haplogroups, with R1a1 reported as the most dominant haplogroup (22.22%) found in the Central Indian population by Y-chromosome 34 SNP analysis using the Precision ID identity panel. An NGS panel consisting of both Y-SNPs (n = 15,611) and Y-STRs (n = 202) (including slow, moderate, and RM Y-STRs) was developed to individualize close paternal relatives worldwide by analyzing the 1443 Y-haplogroup lineages (Claerhout et al., 2021). The SifaMPS panel developed using the predominant 381 Y SNP provided better haplogroup analysis in the Chinese population, which could be adapted for paternal lineage analysis, forensic genealogy studies, and familial searching (Tao et al., 2021). The development of new assays with potential markers will be applicable in diverse populations that need continuous research and validation studies in different populations with large datasets.

Limitations of SNP analysis in forensics

Applications of NGS-based SNP analysis still pose several cons or technical issues in forensic sample analysis. The biallelic nature of most SNPs reduces their discrimination

power per locus in comparison to polymorphic STRs. However, this limitation is easily compensated with a much larger set of SNPs with the use of NGS technique to achieve the same discrimination power as imparted with the core STR loci (Chakraborty et al., 1999). Data interpretation has become difficult as more loci and amplification products are obtained. Locus dropout is another challenge faced with an increase in the number of loci in comparison to conventional SNP genotyping. As a result, MPS-based SNP panels with a larger number of markers may be prone to problems due to low integrity and low copy numbers in the input DNA template. An increase in the number of loci means more peak signals and the likelihood of more artifacts, demanding trained manpower or experts for NGS DNA profile analysis in forensic laboratories. Some issues are being faced with NGS-based SNP panels, such as the low validation rate of some multiplexing genotyping array platforms and the lack of availability of population-specific allele information, which is important for global applicability with the high sensitivity of a panel. The major technical issue in SNP genotyping through many NGS platforms is majorly due to the sequencing of homopolymeric regions (Loman et al., 2015). Some SNP loci, *rs143232,* and *rs1031825*, located inside the homopolymer stretches showed a problem of allelic imbalance on specifically nanopore sequencer, but issues with the Ion Torrent platform have been reported (Loman et al., 2015). However, the high sensitivity of NGS technology could establish a match accuracy of 99.9% while comparing an unknown sample against a database of reference sequences. Nevertheless, this technique seems incongruous in forensic case matching due to the unavailability of DNA sequence databases of criminals. Genetic profiling data available in the forensic DNA database is stored in the form of CE-based STR typing. Hence, constant efforts are needed in the forensic fraternity to make an SNP a replaceable tool of STR in the foreseeable future.

Conclusions

The discovery of SNPs apparently made a breakthrough with the development of NGS technologies. Application of the NGS-based SNP genotyping technique is advantageous in analyzing compromised forensic samples, such as highly degraded and mixed samples, complex kinship and paternity analysis in sexual assault cases, BGA, forensic phenotyping, and identification for missing individuals and unidentified human remains by providing investigative clues without a suspect. Despite the capability of providing sufficient data on loci across the genome, NGS technology needs to work on low template library preparation, error rate, and issues with NGS data processing and mining so that it becomes a routine application in forensic science. Except for human identification via SNP genotyping, NGS is being dominated in other research disciplines, including body fluid identification, identification of monozygotic twins, the age of the perpetrator through DNA methylation techniques, nonhuman species identification, and microbial forensics.

References

Albastaki, A., Naji, M., Lootah, R., Almheiri, R., Almulla, H., Alreyami, A., Almarri, I., Yu, Z., Zhang, Y., Shu, C., & Alghafri, R. (2022). Design and development of novel single multiplex system incorporating 26 rapidly mutating Y-STRs; 26 RM Yplex. *Electrophoresis, 43*(18−19), 1911−1919.

Al-Nbaheen, M. S. (2018). Analysis of Downs syndrome with molecular techniques for future diagnoses. *Saudi Journal of Biological Sciences, 25*(3), 558−562.

Al-Snan, N. R., Shabbir, S., Baksh, S. S., AlQerainees, M., Haidar, M., Messaoudi, S. A., & Bakhiet, M. (2021). Population genetics of 30 insertion/deletion polymorphisms in the Bahraini population. *Scientific Reports, 11*, 6843.

Amorim, A., & Pereira, L. (2005). Pros and cons in the use of SNPs in forensic kinship investigation: A comparative analysis with STRs. *Forensic Science International, 150*(1), 17−21.

Andersen, J. D., Johansen, P., Wulf, H. C., Petersen, B., Børsting, C., & Morling, N. (2011). Genetic variants and skin colour in Danes. *Forensic Science International Genetics, 3*, e153−e154.

Andrade, E. S., Fracasso, N. C. A., Strazza Júnior, P. S., Simões, A. L., & Mendes-Junior, C. T. (2017). Associations of OCA2 - HERC2 SNPs and haplotypes with human pigmentation characteristics in the Brazilian population. *Legal Medicine, 24*, 78−83.

Avent, I., Kinnane, A. G., Jones, N., Petermann, I., Daniel, R., Gahan, M. E., & McNevin, D. (2019). The QIAGEN 140-locus single-nucleotide polymorphism (SNP) panel for forensic identification using massively parallel sequencing (MPS): An evaluation and a direct-to-PCR trial. *International Journal of Legal Medicine, 133*(3), 677−688.

Balanovsky, O., Petrushenko, V., Gorin, I., Kagazezheva, Z., Markina, N. V., Kostryukova, E. S., Leybova, A. N., Maurer, A. M., & Balanovska, E. V. (2019). The accuracy of predicting eye and hair pigmentation based on genetic markers in Russian populations. *Bulletin of Russian State Medical University*, 23−38.

Barany, F. (1991). The ligase chain reaction in a PCR world. *PCR Methods & Applications, 1*, 5−16.

Bardan, F., Higgins, D., & Austin, J. J. (2023). A custom hybridisation enrichment forensic intelligence panel to infer biogeographic ancestry, hair and eye colour, and Y chromosome lineage. *Forensic Science International Genetics, 63*, 102822.

Batnyam, N., Gantulga, A., & Oh, S. (2013). An efficient classification for single nucleotide polymorphism (SNP) dataset. In R. Lee (Ed.), *Computer and information science. Studies in computational intelligence* (Vol 493). Heidelberg: Springer.

Bauchet, M., McEvoy, B., Pearson, L. N., Quillen, E. E., Sarkisian, T., Hovhannesyan, K., Deka, R., Bradley, D. G., & Shriver, M. D. (2007). Measuring European population stratification with microarray genotype data. *The American Journal of Human Genetics, 80*(5), 948−956.

Bennett., L., Oldoni, F., Long, K., Cisana, S., Madella, K., Wootton, S., Chang, J., Hasegawa, R., Lagacé, R., Kidd, K. K., & Podini, D. (2019). Mixture deconvolution by massively parallel sequencing of microhaplotypes. *International Journal of Legal Medicine, 133*(3), 719−729.

Bentley, D. R., Balasubramanian, S., Swerdlow, H. P., Smith, G. P., Milton, J., Brown, C. G., Hall, K. P., Evers, D. J., Barnes, C. L., Bignell, H. R., Boutell, J. M., Bryant, J., Carter, R. J., Keira Cheetham, R., Cox, A. J., Ellis, D. J., Flatbush, M. R., Gormley, N. A., Humphray, S. J., Irving, L. J., Karbelashvili, M. S., … Smith, A. J. (2008). Accurate whole human genome sequencing using reversible terminator chemistry. *Nature, 456*, 53−59.

Biesecker, L. G., Bailey-Wilson, J. E., Ballantyne, J., Baum, H., Bieber, F. R., Brenner, C., Budowle, B., Butler, J. M., Carmody, G., Conneally, P. M., Duceman, B., Eisenberg, A., Forman, L., Kidd, K. K., Leclair, B., Niezgoda, S., Parsons, T. J., Pugh, E., Shaler, R., Sherry, S. T., & Walsh, A. (2005). DNA identifications after the 9/11 world trade center attack. *Science, 310*, 1122−1123.

Boehringer, S., Van Der Lijn, F., Liu, F., Günther, M., Sinigerova, S., Nowak, S., et al. (2011). Genetic determination of human facial morphology: Links between cleft-lips and normal variation. *European Journal of Human Genetics, 19*, 1192.

Børsting, C., & Morling, N. (2011). Mutations and/or close relatives? Six case work examples where 49 autosomal SNPs were used as supplementary markers. *Forensic Science International Genetics, 5*, 236−241.

Bragg, L. M., Stone, G., Butler, M. K., Hugenholtz, P., & Tyson, G. W. (2013). Shining a light on dark sequencing: Characterising errors in ion torrent PGM data. *PLoS Computational Biology, 9*, e1003031.

Branicki, W., Szczerbińska, A., Brudnik, U., Wolańska-Nowak, P., & Kupiec, T. (2008). The OCA2 gene as a marker for eye colour prediction. *Forensic Science International: Genetics, 1*(1), 536–537. https://doi.org/10.1016/j.fsigss.2007.10.062

Braun, A., Little, D. P., & Koster, H. (1997). Detecting CFTR gene mutations by using primer oligo base extension and mass spectrometry. *Clinical Chemistry, 43*, 1151–1158.

Brinkley, J. F., Fisher, S., Harris, M. P., Holmes, G., Hooper, J. E., Jabs, E. W., Jones, K. L., Kesselman, C., Klein, O. D., Maas, R. L., Marazita, M. L., Selleri, L., Spritz, R. A., van Bakel, H., Visel, A., Williams, T. J., Wysocka, J., Chai, Y., & FaceBase Consortium. (2016). The FaceBase consortium: A comprehensive resource for craniofacial researchers. *Development, 143*, 2677–2688.

Budowle, B. (2004). SNP typing strategies. *Forensic Science International, 146*, 139–142. https://doi.org/10.1016/j.forsciint.2004.09.042

Budowle, B., & Baechtel, F. S. (1990). Modifications to improve the effectiveness of restriction fragment length polymorphism typing. *Applied and Theoretical Electrophoresis, 1*(4), 181–187.

Budowle, B., Shea, B., Niezgoda, S., & Chakraborty, R. (2001). CODIS STR loci data from 41 sample populations. *Journal of Forensic Sciences, 46*, 453–489.

Budowle, B., Lindsey, J. A., DeCou, J. A., Koons, B. W., Giusti, A. M., & Comey, C. T. (1995). Validation and population studies of the loci LDLR, GYPA, HBGG, D7S8, and Gc (PM loci), and HLA-DQα using a multiplex amplification and typing procedure. *Journal of Forensic Sciences, 40*, 45–54.

Budowle, B., Planz, J., Campbell, R., & Eisenberg, A. (2004). SNPs and microarray technology in forensic genetics: Development and application to mitochondrial DNA. *Forensic Science Review, 16*, 22–36.

Butler, J. M. (2011). *Advanced topics in forensic DNA typing: Methodology* (pp. 347–362). Academic Press.

Cao, M., Shi, J., Wang, Ji, Hong, J., Cui, B., & Ning, G. (2015). Analysis of human triallelic SNPs by next-generation sequencing. *Annals of Human Genetics, 79*, 275–328.

Carrasco, P., Inostroza, C., Didier, M., Godoy, M., Holt, C. L., Tabak, J., & Loftus, A. (2020). Optimizing DNA recovery and forensic typing of degraded blood and dental remains using a specialized extraction method, comprehensive qPCR sample characterization, and massively parallel sequencing. *International Journal of Legal Medicine, 134*(1), 7.

Carratto, T. M. T., Moraes, V. M. S., Recalde, T. S. F., Oliveira, M. L. G., & Teixeira Mendes-Junior, C. (2022). Applications of massively parallel sequencing in forensic genetics. *Genetics and Molecular Biology, 45*(3 Suppl. 1), e20220077.

Chaitanya, L., Breslin, K., Zuñiga, S., Wirken, L., Pośpiech, E., KuklaBartoszek, M., Sijen, T., Knijff, P., Liu, F., Branicki, W., Kayser, M., & Walsh, S. (2018). The HIrisPlex-S system for eye, hair and skin colour prediction from DNA: Introduction and forensic developmental validation. *Forensic Science International Genetics, 35*, 123–135.

Chakraborty, R., Stivers, D. N., Su, B., Zhong, Y., & Budowle, B. (1999). The utility of short tandem repeat loci beyond human identification: Implications for development of new DNA typing systems. *Electrophoresis, 20*(8), 1682–1696.

Chen, X., Livak, K. J., & Kwok, P. Y. (1998). A homogeneous, ligase-mediated DNA diagnostic test. *Genome Research, 8*(5), 549–556.

Chen, P., Deng, C., Li, Z., Pu, Y., Yang, J., Yu, Y., Li, K., Li, D., Liang, W., Zhang, L., & Chen, F. (2019). A microhaplotypes panel for massively parallel sequencing analysis of DNA mixtures. *Forensic Science International: Genetics, 40*, 140–149.

Chen, X., Levine, L., & Kwok, P. Y. (1999). Fluorescence polarization in homogeneous nucleic acid analysis. *Genome Research, 9*(5), 492–498.

Chen, R. H., Zhao, X. Y., Ma, K., Li, H., Cao, Y., Cao, Y., & Liu, W. (2020). Separation of SNP profiles from DNA mixtures with two contributors via massively parallel sequencing technology. *Australian Journal of Forensic Sciences, 52*(5), 537–546.

Cho, S., Shin, K. J., Bae, S. J., Kwon, Y. L., & Lee, S. D. (2020). Improved STR analysis of degraded DNA from human skeletal remains through in-house MPS-STR panel. *Electrophoresis, 41*(18–19), 1600–1605.

Churchill, J. D., Schmedes, S. E., King, J. L., & Budowle, B. (2016). Evaluation of the illumina beta version ForenSeq DNA signature prep kit for use in genetic profiling. *Forensic Science International Genetics, 20*, 20–29.

Claerhout, S., Verstraete, P., Warnez, L., Vanpaemel, S., Larmuseau, M., & Decorte, R. (2021). The first Y-chromosome sequencing tool typing a large number of Y-SNPs and Y-STRs to unravel worldwide human population genetics. *PLoS Genetics, 17*(9), e1009758. https://doi.org/10.1371/journal.pgen.1009758

Clegg, R. M. (1992). Fluorescence resonance energy transfer and nucleic acids. *Methods in Enzymology, 221*, 353–388.

Comey, C. T., & Budowle, B. (1991). Validation studies on the analysis of the HLA DQ alpha locus using the polymerase chain reaction. *Journal of Forensic Sciences, 36*(6), 1633–1648.

Cornelius, S., Gansemans, Y., Vander Plaetsen, A. S., Weymaere, J., Willems, S., Deforce, D., & Van Nieuwerburgh, F. (2019). Forensic tri-allelic SNP genotyping using nanopore sequencing. *Forensic Science International Genetics, 38*, 204–210.

Dash, H. R., Avila, E., Jena, S. R., Kaitholia, K., Agarwal, R., Alho, C. S., Srivastava, A., & Singh, A. K. (2022). Forensic characterization of 124 SNPs in the central Indian population using precision ID identity panel through next-generation sequencing. *International Journal of Legal Medicine, 136*(2), 465–473.

Dash, H. R., Kaitholia, K., Kumawat, R. K., Singh, A. K., Shrivastava, P., Chaubey, G., & Das, S. (2021). Sequence variations, flanking region mutations, and allele frequency at 31 autosomal STRs in the central Indian population by next generation sequencing (NGS). *Scientific Reports, 11*, 23238.

Dash, H. R., Vajpayee, K., Shukla, R., Srivastava, A., Shrivastava, P., & Das, S. (2022). Sequence-based assessment of expediency of tri-, tetra-, and penta-nucleotides repeat autosomal STR markers in the central Indian population using next generation sequencing (NGS). *Meta Gene, 31*, 100983.

de Knijff, P. (2022). On the forensic use of Y-chromosome polymorphisms. *Genes, 13*(5), 898.

Deka, R., Jin, L., Shriver, M. D., Yu, L. M., DeCroo, S., Hundrieser, J., Bunker, C. H., Ferrell, R. E., & Chakraborty, R. (1995). Population genetics of dinucleotide (dC−dA)*n*. (dG−dT)*n*polymorphisms in world populations. *The American Journal of Human Genetics, 56*, 461–474.

Devesse, L., Ballard, D., Davenport, L., Riethorst, I., Mason-Buck, G., & Syndercombe Court, D. (2018). Concordance of the ForenSeq system and characterization of sequence-specific autosomal STR alleles across two major population groups. *Forensic Science International Genetics, 34*, 57–61.

Divne, A. M., & Allen, M. (2005). A DNA microarray system for forensic SNP analysis. *Forensic Science International Genetics, 154*, 111–121.

Duffy, D. L., Montgomery, G. W., Chen, W., Zhao, Z. Z., Le, L., James, M. R., Hayward, N. K., Martin, N. G., & Sturm, R. A. (2007). A three-single-nucleotide polymorphism haplotype in intron 1 of OCA2 explains most human eye-color variation. *The American Journal of Human Genetics, 80*(2), 241–252.

Fan, G., Pan, L., Tang, P., Zhou, Y., Liu, M., & Luo, X. (2021). Technical note: Developmental validation of a novel 41-plex Y-STR system for the direct amplification of reference samples. *International Journal of Legal Medicine, 135*(2), 409–419.

Frudakis, T. N. (2008). *Molecular photofitting: Predicting ancestry and phenotype from DNA*. Amsterdam,Netherlands: Academic Press Publishers (Elsevier).

Gao, Z., Chen, X., Zhao, Y., Zhao, X., Zhang, S., Yang, Y., Wang, Y., & Zhang, J. (2018). Forensic genetic informativeness of an SNP panel consisting of 19 multi-allelic SNPs. *Forensic Science International Genetics, 34*, 49–56.

Gettings, K. B., Lai, R., Johnson, J. L., Peck, M. A., Hart, J. A., Gordish-Dressman, H., Schanfield, M. S., & Podini, D. S. (2014). A 50-SNP assay for biogeographic ancestry and phenotype prediction in the U.S. population. *Forensic Science International Genetics, 8*(1), 101–108.

Gill, P. (2001). An assessment of the utility of single nucleotide polymorphisms (SNPs) for forensic purposes. *International Journal of Legal Medicine, 114*, 204–210.

Giusti, A. M., & Budowle, B. (1995). A chemiluminescence-based detection system for human DNA quantitation and restriction fragment length polymorphism (RFLP) analysis. *Applied and Theoretical Electrophoresis, 5*(2), 89–98.

González-Herrera, L. J., García-Aceves, M. E., Domínguez-Cruz, M. D., López-González, P. N., Sosa-Escalante, J. E., & Rangel-Villalobos, H. (2020). A four-step mutation at D22S1045 in one complex

paternity case when the brother of the alleged father hypothesis is evaluated. *International Journal of Legal Medicine, 134*(5), 1647–1652. https://doi.org/10.1007/s00414-020-02312-1

Goodwin, S., McPherson, J. D., & McCombie, W. R. (2016). Coming of age: Ten years of next-generation sequencing technologies. *Nature Reviews Genetics, 17*, 333–351.

Griffin, T. J., Hall, J. G., Prudent, J. R., & Smith, L. M. (1999). Direct genetic analysis by matrix-assisted laser desorption/ionization mass spectrometry. *Proceedings of the National Academy of Sciences of the United States of America, 96*(11), 6301–6306.

Grossman, P. D., Bloch, W., Brinson, E., Chang, C. C., Eggerding, F. A., Fung, S., Iovannisci, D. M., Woo, S., & Winn-Deen, E. S. (1994). High-density multiplex detection of nucleic acid sequences: Oligonucleotide ligation assay and sequence-coded separation. *Nucleic Acids Research, 22*(21), 4527–4534.

Gyllensten, U. B., & Erlich, H. A. (1988). Generation of single-stranded DNA by the polymerase chain reaction and its application to direct sequencing of the HLA-DQA locus. *Proceedings of the National Academy of Sciences of the United States of America, 85*(20), 7652–7656.

Haff, L. A., & Smirnov, I. P. (1997). Single-nucleotide polymorphism identification assays using a thermostable DNA polymerase and delayed extraction MALDI-TOF mass spectrometry. *Genome Research, 7*, 378–388.

Harris, K., & Pritchard, J. K. (2017). Rapid evolution of the human mutation spectrum. *Elife, 6*, e24284. pii.

Holden, A. L. (2002). The SNP consortium: Summary of a private consortium effort to develop an applied map of the human genome. *Biotechniques, 32*, S22–S26.

Holland, P. M., Abramson, R. D., Watson, R., & Gelfand, D. H. (1991). Detection of specific polymerase chain reaction product by utilizing the 5'–3' exonuclease activity of thermus aquaticus DNA polymerase. *Proceedings of the National Academy of Sciences of the United States of America, 88*, 7276–7280.

Holland, M. M., McQuillan, M. R., & O'Hanlon, K. A. (2011). Second generation sequencing allows for mtDNA mixture deconvolution and high resolution detection of heteroplasmy. *Croatian Medical Journal, 52*(3), 299–313.

Hsu, T. M., Law, S. M., Duan, S., Neri, B. P., & Kwok, P. Y. (2001). Genotyping single-nucleotide polymorphisms by the invader assay with dual-color fluorescence polarization detection. *Clinical Chemistry, 47*(8), 1373–1377.

Hwa, H. L., Wu, M. Y., Lee, J. C., Yin, H. I., Hsu, P. M., Li, S. F., Hwu, W. L., & Su, C. W. (2020). Analysis of nondegraded and degraded DNA mixtures of close relatives using massively parallel sequencing. *Legal Medicine, 42*, 101631.

Jäger, A. C., Alvarez, M. L., Davis, C. P., Guzmán, E., Han, Y., Way, L., Walichiewicz, P., Silva, D., Pham, N., Caves, G., Bruand, J., Schlesinger, F., Pond, S. J. K., Varlaro, J., Stephens, K. M., & Holt, C. L. (2017). Developmental validation of the MiSeq FGx Forensic Genomics System for Targeted Next Generation Sequencing in Forensic DNA Casework and Database Laboratories. *Forensic Science International: Genetics, 28*, 52–70. https://doi.org/10.1016/j.fsigen.2017.01.011

Jeffreys, A. J., Wilson, V., & Thein, S. L. (1985). Hypervariable 'Minisatellite' regions in human DNA. *Nature, 314*(6006), 67–73, 7-13.

Kaiser, M. W., Lyamicheva, N., Ma, W., Miller, C., Neri, B., Fors, L., & Lyamichev, V. I. (1999). A comparison of eubacterial and archaeal structure-specific 5'-exonucleases. *Journal of Biological Chemistry, 274*(30), 21387–21394.

Kayser, M. (2015). Forensic DNA Phenotyping: Predicting human appearance from crime scene material for investigative purposes. *Forensic Science International Genetics, 18*, 33–48. https://doi.org/10.1016/j.fsigen.2015.02.003

Kayser, M., Liu, F., Cecile, J. W., Rivadeneira, F., Lao, O., van Duijn, K., Vermeulen, M., Arp, P., Jhamai, M. M., van Ijcken, W. F., den Dunnen, J. T., Heath, S., Zelenika, D., Despriet, D. D., Klaver, C. C., Vingerling, J. R., de Jong, P. T., Hofman, A., Aulchenko, Y. S., Uitterlinden, A. G., & van Duijn, C. M. (2008). Three genome-wide association studies and a linkage analysis identify HERC2 as a human Iris color gene. *The American Journal of Human Genetics, 82*(2), 411–423.

Kidd, K. K., Pakstis, A. J., Speed, W. C., Grigorenko, E. L., Kajuna, S. L., Karoma, N. J., Kungulilo, S., Kim, J. J., Lu, R. B., Odunsi, A., Okonofua, F., Parnas, J., Schulz, L. O., Zhukova, O. V., &

Kidd, J. R. (2006). Developing a SNP panel for forensic identification of individuals. *Forensic Science International, 164*, 20–32.

Kircher, M., & Kelso, J. (2010). High-throughput DNA sequencing-concepts and limitations. *BioEssays, 32*(6), 524–536.

Kosoy, R, Nassir, R., Tian, C., White, P. A., Butler, L. M., Silva, G., Kittles, R., Alarcon-Riquelme, M. E., Gregersen, P. K., Belmont, J. W., De La Vega, F. M., & Seldin, M. F. (2009). Ancestry informative marker sets for determining continental origin and admixture proportions in common populations in America. *Human Mutation, 30*, 69–78.

Kuppuswamy, M. N., Hoffmann, J. W., Kasper, C. K., Spitzer, S. G., Groce, S. L., & Bajaj, S. P. (1991). Single nucleotide primer extension to detect genetic diseases: Experimental application to hemophilia B (factor IX) and cystic fibrosis genes. *Proceedings of the National Academy of Sciences of the United States of America, 88*(4), 1143–1147.

Landegren, U., Kaiser, R., Sanders, J., & Hood, L. (1988). A ligase-mediated gene detection technique. *Science, 241*(4869), 1077–1080.

Lao, O., van Duijn, K., Kersbergen, P., de Knijff, P., & Kayser, M. (2006). Proportioning whole-genome single-nucleotide-polymorphism diversity for the identification of geographic population structure and genetic ancestry. *The American Journal of Human Genetics, 78*(4), 680–690.

Lareu, M., Puente, J., Sobrino, B., Quintans, B., Brion, M., & Carracedo, A. (2001). The use of the LightCycler for the detection of Y chromosome SNPs. *Forensic Science International, 118*, 163–168.

Liu, Y., Guo, Y., Jin, X., Mei, S., Xie, T., Lan, Q., Fang, Y., & Zhu, B. (2020). Developmental validation study of a 24-plex Y-STR direct amplification system for forensic application. *International Journal of Legal Medicine, 134*(3), 945–951.

Liu, F., van Duijn, K., Vingerling, J. R., Hofman, A., Uitterlinden, A. G., Janssens, A. C., & Kayser, M. (2009). Eye color and the prediction of complex phenotypes from genotypes. *Current Biology, 19*, R192–R193.

Loman, N. J., Quick, J., & Simpson, J. T. (2015). A complete bacterial genome assembled de novo using only nanopore sequencing data. *Nature Methods, 12*, 733–735.

Lorenzo, F., Huff, C., Myllymäki, M., Olenchock, B., Swierczek, S., Tashi, T., GordeuK, V., Wuren, T., Ri-Li, G., McClain, D. M, Khan, T. M., Koul, P. A., Guchhait, P., Salama, M. E., Xing, J., Semenza, G. L., Liberzon, E., Wilson, A., Simonson, T. S., Jorde, L. B., & Prchal, J. T. (2014). A genetic mechanism for Tibetan high-altitude adaptation. *Nature Genetics, 46*, 951–956.

Lyamichev, V., Mast, A. L., Hall, J. G., Prudent, J. R., Kaiser, M. W., Takova, T., Kwiatkowski, R. W., Sander, T. J., de Arruda, M., Arco, D. A., Neri, B. P., & Brow, M. A. (1999). Polymorphism identification and quantitative detection of genomic DNA by invasive cleavage of oligonucleotide probes. *Nature Biotechnology, 17*(3), 292–296.

Malhotra, S., Singh, S., & Sarkar, S. (2018). Whole genome variant analysis in three ethnically diverse Indians. *Genes Genomics, 40*(5), 497–510.

Martínez-Arias, R., Calafell, F., Mateu, E., Comas, D., Andrés, A., & Bertranpetit, J. (2001). Sequence variability of a human pseudogene. *Genome Research, 11*(6), 1071–1085.

Mehmood, T., Liland, K. H., Snipen, L., & Sæbø, S. (2012). A review of variable selection methods in partial least squares regression. *Chemometrics and Intelligent Laboratory Systems, 118*, 62–69.

Mei, R., Galipeau, P. C., Prass, C., Berno, A., Ghandour, G., Pati, N., Wolff, R. K., Chee, M. S., Reid, B. J., & Lockhart, D. J. (2000). Genome-wide detection of allelic imbalance using human SNPs and high-density DNA arrays. *Genome Research, 10*, 1126–1137.

Mein, C. A., Barratt, B. J., Dunn, M. G., Siegmund, T., Smith, A. N., Esposito, L., Nutland, S., Stevens, H. E., Wilson, A. J., Phillips, M. S., Jarvis, N., Law, S., de Arruda, M., & Todd, J. A. (2000). Evaluation of single nucleotide polymorphism typing with invader on PCR amplicons and its automation. *Genome Research, 10*(3), 330–343.

Mitchelson, K. R. (2003). The use of capillary electrophoresis for DNA polymorphism analysis. *Molecular Biotechnology*, 41–68.

Momota, F., Tsuji, A., Ishiko, A., & Ikeda, N. (2021). Examination of the usefulness of next-generation sequencing in mixed DNA samples. *Legal Medicine, 51*, 101874.

Nachman, M. W., & Crowell, S. L. (2000). Estimate of the mutation rate per nucleotide in humans. *Genetics, 156*, 297−304.

Nei, M. (1987). *Molecular evolutionary genetics*. New York: Columbia University Press.

Ng, S., Buckingham, K., Lee, C., Bigham, A. W., Tabor, H. K., Dent, K. M., Huff, C. D., Shannon, P. T., Jabs, E. W., Nickerson, D. A., Shendure, J., & Bamshad, M. J. (2010). Exome sequencing identifies the cause of a Mendelian disorder. *Nature Genetics, 42*, 30−35.

Ng, S. B., Turner, E. H., Robertson, P. D., Flygare, S. D., Bigham, A. W., Lee, C., Shaffer, T., Wong, M., Bhattacharjee, A., Eichler, E. E, Bamshad, M., Nickerson, D. A., & Shendure, J. (2009). Targeted capture and massively parallel sequencing of 12 human exomes. *Nature, 461*(7261), 272−276.

Oostdik, K., Lenz, K., Nye, J., Schelling, K., Yet, D., Bruski, S., Strong, J., Buchanan, C., Sutton, J., Linner, J., Frazier, N., Young, H., Matthies, L., Sage, A., Hahn, J., Wells, R., Williams, N., Price, M., Koehler, J., Staples, M., & Storts, D. R. (2014). Developmental validation of the PowerPlex fusion system for analysis of casework and reference samples: A 24-locus multiplex for new database standards. *Forensic Science International Genetics, 12*, 69−76.

Palencia-Madrid, L., Xavier, C., de la Puente, M., Hohoff, C., Phillips, C., Kayser, M., & Parson, W. (2020). Evaluation of the VISAGE basic tool for appearance and ancestry prediction using PowerSeq chemistry on the MiSeq FGx system. *Genes, 11*(6), 708.

Pastinen, T., Raitio, M., Lindroos, K., Tainola, P., Peltonen, L., & Syvänen, A. C. (2000). A system for specific, high-throughput genotyping by allele-specific primer extension on microarrays. *Genome Research, 10*(7), 1031−1042.

Pereira, V., Freire-Aradas, A., Ballard, D., Børsting, C., Diez, V., Pruszkowska-Przybylska, P., Ribeiro, J., Achakzai, N. M., Aliferi, A., Bulbul, O., Carceles, M. D. P., Triki-Fendri, S., Rebai, A., Court, D. S., Morling, N., Lareu, M. V., Carracedo, A., EUROFORGEN-NoE Consortium, & Phillips, C. (2019). Development and validation of the Euroforgen name (North African and Middle Eastern) ancestry panel. *Forensic Science International Genetics, 42*, 260−267.

Phillips, C., Amigo, J., Tillmar, A. O., Peck, M. A., de la Puente, M., Ruiz-Ramírez, J., Bittner, F., Idrizbegović, Š, Wang, Y., Parsons, T. J., & Lareu, M. V. (2020). A compilation of tri-allelic SNPs from 1000 Genomes and use of the most polymorphic loci for a large-scale human identification panel. *Forensic Science International Genetics, 46*, 102232.

Petrovick, M. S., Boettcher, T., Fremont-Smith, P., Peragallo, C., Ricke, D. O., Watkins, J., & Schwoebel, E. (2020). Analysis of complex DNA mixtures using massively parallel sequencing of SNPs with low minor allele frequencies. *Forensic Science International: Genetics, 46*, 102234.

Phillips, C., Amigo, J., Carracedo, A., & Lareu, M. V. (2015). Tetra-allelic SNPs: Informative forensic markers compiled from public whole-genome sequence data. *Forensic Science International Genetics, 19*, 100−106.

Phillips, C., Freire Aradas, A., Kriegel, A. K., Fondevila, M., Bulbul, O., Santos, C., Serrulla Rech, F., Perez Carceles, M. D., Carracedo, A., Schneider, P. M., & Lareu, M. V. (2013). Eurasiaplex: A forensic SNP assay for differentiating European and South Asian ancestries. *Forensic Science International Genetics, 7*(3), 359−366.

Phillips, C., Manzo, L., de la Puente, M., Fondevila, M., & Lareu, M. V. (2020). The MASTiFF panel-a versatile multiple-allele SNP test for forensics. *International Journal of Legal Medicine, 134*(2), 441−450.

Phillips, C., Prieto, L., Fondevila, M., Salas, A., Gómez-Tato, A., Álvarez-Dios, J., Alonso, A., Blanco-Verea, A., Brión, M., Montesino, M., Carracedo, A., & Victoria Lareu, M. (2009). Ancestry analysis in the 11-M madrid bomb attack investigation. *PLoS One, 4*(8), e6583.

Pilli, E., Morelli, S., Poggiali, B., & Alladio, E. (2023). Biogeographical ancestry, variable selection, and PLS-DA method: A new panel to assess ancestry in forensic samples via MPS technology. *Forensic Science International Genetics, 62*, 102806.

Qin, D. (2019). Next-generation sequencing and its clinical application. *Cancer Biology Medicine, 16*(1), 4−10.

Qu, S., Zhu, J., Wang, Y., Yin, L., Lv, M., Wang, L., Jian, H., Tan, Y., Zhang, R., Liu, Y., Li, F., Huang, S., Liang, W., & Zhang, L. (2019). Establishing a second-tier panel of 18 ancestry informative markers to improve ancestry distinctions among Asian populations. *Forensic Science International Genetics, 41*, 159−167.

Quintáns, B., Alvarez-Iglesias, V., Salas, A., Phillips, C., Lareu, M. V., & Carracedo, A. (2004). Typing of mitochondrial DNA coding region SNPs of forensic and anthropological interest using SNaPshot minisequencing. *Forensic Science International, 140*(2–3), 251–257.

Ricke, D. O., Watkins, J., Fremont-Smith, P., Petrovick, M. S., Boettcher, T., & Schwoebel, E. (2021). TranslucentID: Analysis of complex DNA SNP mixtures with large numbers of donors. *Australian Journal of Forensic Sciences, 53*(4), 378–389.

Riddell, J., Basu Mallick, C., Jacobs, G. S., Schoenebeck, J. J., & Headon, D. J. (2020). Characterisation of a second gain of function EDAR variant, encoding EDAR380R, in East Asia. *European Journal of Human Genetics, 28*(12), 1694–1702.

Rockenbauer, E., Hansen, S., Mikkelsen, M., Børsting, C., & Morling, N. (2014). Characterization of mutations and sequence variants in the D21S11 locus by next generation sequencing. *Forensic Science and International Genetics, 8*(1), 68–72.

Ronaghi, M., Uhlén, M., & Nyrén, P. (1998). A sequencing method based on real-time pyrophosphate. *Science, 281*(5375), 363–365.

Ruiz, Y., Phillips, C., Gomez-Tato, A., Alvarez-Dios, J., Casares de Cal, M., Cruz, R., Maroñas, O., Söchtig, J., Fondevila, M., Rodriguez-Cid, M. J., Carracedo, A., & Lareu, M. V. (2013). Further development of forensic eye color predictive tests. *Forensic Science International Genetics,* 7(1), 28–40.

Sachidanandam, R., Weissman, D., Schmidt, S. C., Kakol, J. M., & Willey, D. L. (2001). A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature, 409*, 928–933.

Saiki, R. K., Scharf, S., Faloona, F., Mullis, K. B., Horn, G. T., Erlich, H. A., & Arnheim, N. (1985). Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science, 230*(4732), 1350–1354.

Sanchez, J., Børsting, C., Hallenberg, C., Buchard, A., Hernandez, A., Sobrino, B., Brion, M., & Carracedo, A. (2005). SNPs in forensic genetics: A review on SNP typing methodologies. *Forensic Science International, 154*, 181–194.

Sanchez, J. J., Phillips, C., Børsting, C., Balogh, K., Bogus, M., Fondevila, M., Harrison, C. D., Musgrave-Brown, E., Salas, A., Syndercombe-Court, D., Schneider, P. M., Carracedo, A., & Morling, N. (2006). A multiplex assay with 52 single nucleotide polymorphisms for human identification. *Electrophoresis, 27*(9), 1713–1724.

Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America, 74*(12), 5463–5467.

Santos, C., Phillips, C., Fondevila, M., Daniel, R., van Oorschot, R. A. H., Burchard, E. G., Schanfield, M. S., Souto, L., Uacyisrael, J., Via, M., Carracedo, Á., & Lareu, M. V. (2016). Pacifiplex: An ancestry-informative SNP panel centred on Australia and the pacific region. *Forensic Science International Genetics, 20*, 71–80.

Senst, A., Caliebe, A., Eva Scheurer, M. D., & Iris Schulz, P. (2022). Validation and beyond: Next generation sequencing of forensic casework samples including challenging tissue samples from altered human corpses using the MiSeq FGx system. *Journal of Forensic Sciences, 67*, 1382–1398.

Seo, S. B., King, J. L., Warshauer, D. H., Davis, C. P., Ge, J., & Budowle, B. (2013). Single nucleotide polymorphism typing with massively parallel sequencing for human identification. *International Journal of Legal Medicine, 127*, 1079–1086.

Shang, L., Ding, G., Mo, X., Sun, J., Sun, H., Yu, Z., & Li, W. (2021). A novel multiplex of 12 multicopy Y-STRs for forensic application. *Journal of Forensic Sciences, 66*(5), 1901–1907.

Shriver, M. D., Smith, M. W., Jin, L., Marcini, A., Akey, J. M., Deka, R., & Ferrell, R. E. (1997). Ethnic-affiliation estimation by use of population-specific DNA markers. *The American Journal of Human Genetics, 60*(4), 957–964.

Silvery, J., Ganschow, S., Wiegand, P., & Shriver Shriver Tiemann, C. (2020). Developmental validation of the monSTR identity panel, a forensic STR multiplex assay for massively parallel sequencing. *Forensic Science International Genetics., 46*, 102236.

Sobiah, R., Syeda, R. H., Zunaira, E., Nageen, Z., Maria, K., Syeda, A. Z., Shahana, S. M., Akifa, M., Abdul, J., & Muhammad, R. K. (2018). Implications of targeted next generation sequencing in forensic science. *Journal of Forensic Research, 9*, 416.

Sobrino, B., Brion, M., & Carracedo, A. (2005). SNPs in forensic genetics: a review on SNP typing methodologies. *Forensic Science International, 154*, 181–194. https://doi.org/10.1016/j.forsciint.2004.10.020

Syndercombe Court, D. (2021). The Y chromosome and its use in forensic DNA analysis. *Emerging Top Life Science, 5*(3), 427–441.

The 1000 Genomes Project Consortium, Auton, A., Brooks, L. D., Durbin, R. M., Garrison, E. P., Kang, H. M., Korbel, J. O., Marchini, J. L., McCarthy, S., McVean, G. A., et al. (2015). A global reference for human genetic variation. *Nature, 526*, 68–74.

Tao, R., Wang, S., Chen, A., Xia, R., Zhang, X., Yang, Q., Qu, Y., Zhang, S., & Li, C. (2021). Parallel sequencing of 87 STR and 294 SNP markers using the prototype of the SifaMPS panel on the MiSeq FGx system. *Forensic Science International Genetics, 52*, 102490.

Tillmar, A., Sturk-Andreaggi, K., Daniels-Higginbotham, J., Thomas, J. T., & Marshall, C. (2021). The FORCE panel: An all-in-one SNP marker set for confirming investigative genetic genealogy leads and for general forensic applications. *Genes, 12*(12), 1968.

Turchi, C., Previdere, C., Bini, C., Carnevali, E., Grignani, P., Manfredi, A., Melchionda, F., Onofri, V., Pelotti, S., Robino, C., Sorçaburu-Ciglieri, S., Tagliabracci, A., & Fattorini, P. (2020). Assessment of the precision Id identity panel kit on challenging forensic samples. *Forensic Science International Genetics, 49*, 102400.

Tyagi, S., & Kramer, F. R. (1996). Molecular beacons: Probes that fluoresce upon hybridization. *Nature Biotechnology, 14*, 303–308.

Valenzuela, R. K., Henderson, M. S., Walsh, M. H., Garrison, N. A., Kelch, J. T., Cohen-Barak, O., Erickson, D. T., John Meaney, F., Bruce Walsh, J., Cheng, K. C., Ito, S., Wakamatsu, K., Frudakis, T., Thomas, M., & Brilliant, M. H. (2010). Predicting phenotype from genotype: Normal pigmentation. *Journal of Forensic Sciences, 55*, 315–322.

Valle-Silva, G. D., Souza, F. D. N., Marcorin, L., Pereira, A. L. E., Carratto, T. M. T., Debortoli, G., Oliveira, M. L. G., Fracasso, N. C. A., Andrade, E. S., Donadi, E. A., Norton, H. L., Parra, E. J., Simões, A., Castelli, E. C., & Mendes-Junior, C. T. (2019). Applicability of the SNPforID 52-plex panel for human identification and ancestry evaluation in a Brazilian population sample by next-generation sequencing. *Forensic Science International Genetics, 40*, 201–209.

van Daal, A. (2008). The genetic basis of human pigmentation. *Forensic Science International Genetics, 1*, 541–543.

van Dijk, E. L., Auger, H., Jaszczyszyn, Y., & Thermes, C. (2014). Ten years of next-generation sequencing technology. *Trends in Genetics, 30*, 418–426.

Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., Holt, R. A., Gocayne, J. D., Amanatides, P., Ballew, R. M., Huson, D. H., Wortman, J. R., Zhang, Q., Kodira, C. D., Zheng, X. H., Chen, L., Skupski, M., …, & Zhu, X. (2001). The sequence of the human genome. *Science, 291*(5507), 1304–1351.

Vernot, B., Zavala, E. I., Gómez-Olivencia, A., Jacobs, Z., Slon, V., Mafessoni, F., Romagné, F., Pearson, A., Petr, M., Sala, N., Pablos, A., Aranburu, A., de Castro, J. M. B., Carbonell, E., Li, B., Krajcarz, M. T., Krivoshapkin, A. I., Kolobova, K. A., Kozlikin, M. B., Shunkov, M. V., … Meyer, M. (2021). Unearthing Neanderthal population history using nuclear and mitochondrial DNA from cave sediments. *Science, 372*(6542), eabf1667, 7.

Wallace, R. B., Shaffer, J., Murphy, R. F., Bonner, J., Hirose, T., & Itakura, K. (1979). Hybridization of synthetic oligodeoxyribonucleotides to phi chi 174 DNA: The effect of single base pair mismatch. *Nucleic Acids Research, 6*(11), 3543–3557.

Walsh, P. S., Fildes, N., Louie, A. S., & Higuchi, R. (1991). Report of the blind trial of the cetus AmpliType HLA DQα forensic deoxyribonucleic acid (DNA) amplification and typing kit. *Journal of Forensic Sciences, 36*, 1551–1556.

Walsh, S., Lindenbergh, A., Zuniga, S. B., Sijen, T., de Knijff, P., Kayser, M., & Ballantyne, K. N. (2011). Developmental validation of the IrisPlex system: Determination of blue and brown iris colour for forensic intelligence. *Forensic Science International Genetics, 5*, 464–471.

Walsh, S., Wollstein, A., Liu, F., Chakravarthy, U., Rahu, M., Seland, J. H., Soubrane, G., Tomazzoli, L., Topouzis, F., Vingerling, J. R., Vioque, J., Fletcher, A. E., Ballantyne, K. N., & Kayser, M. (2012). DNA-based eye colour prediction across Europe with the IrisPlex system. *Forensic Science International Genetics, 6*, 330–340.

Wang, L., Hea, W., Mao, J., Wang, H., Jin, B., Luo, H. B., Liang, W. B., & Zhang, L. (2015). Development of a SNP-STR multiplex for forensic identification. *Forensic Science International Genetics, 5*, e598–e600.

Westen, A. A., Matai, A. S., Laros, J. F., Meiland, H. C., Jasper, M., de Leeuw, W. J., de Knijff, P., & Sijen, T. (2009). Tri-allelic SNP markers enable analysis of mixed and degraded DNA samples. *Forensic Science International Genetics, 3*(4), 233–241.

Wyman, A. R., & White, R. (1980). A highly polymorphic locus in human DNA. *Proceedings of the National Academy of Sciences of the United States of America, 77*(11), 6754–6758.

Xavier, C., de la Puente, M., Phillips, C., Eduardoff, M., Heidegger, A., Mosquera-Miguel, A., Freire-Aradas, A., Lagace, R., Wootton, S., Power, D., Parson, W., Lareu, M. V., & Daniel, R. (2020). Forensic evaluation of the Asia Pacific ancestry-informative MAPlex assay. *Forensic Science International Genetics, 48*, 102344.

Xiang, K., Ouzhuluobu, Peng, Y., Yang, Z., Zhang, X., Cui, C., Zhang, H., Li, M., Zhang, Y., Bianba, Gonggalanzi, Basang, Ciwangsangbu, Wu, T., Chen, H., Shi, H., Qi, X., & Su, B. (2013). Identification of a Tibetan-specific mutation in the hypoxic gene EGLN1 and its contribution to high-altitude adaptation. *Molecular Biology and Evolution, 30*(8), 1889–1898.

Xiong, Z., Dankova, G., Howe, L. J., Lee, M. K., Hysi, P. G., de Jong, M. A., Zhu, G., Adhikari, K., Li, D., Li, Y., Pan, B., Feingold, E., Marazita, M. L., Shaffer, J. R., McAloney, K., Xu, S. H., Jin, L., Wang, S., de Vrij, F. M., Lendemeijer, B., Richmond, S, … Kayser, M. (2019). Novel genetic loci affecting facial shape variation in humans. *Elife, 8*, e49898.

Yang, Y., Muzny, D. M., Reid, J. G., Bainbridge, M. N., Willis, A., Ward, P. A., Braxton, A., Beuten, J., Xia, F., Niu, Z., Hardison, M., Person, R., Bekheirnia, M. R., Leduc, M. S., Kirby, A., Pham, P., Scull, J., Wang, M., Ding, Y., Plon, S. E., … Eng, C. M. (2013). Clinical whole-exome sequencing for the diagnosis of Mendelian disorders. *New England Journal of Medicine, 369*(16), 1502–1511.

Yun, L., Gu, Y., Rajeevan, H., & Kidd, K. K. (2014). Application of six IrisPlex SNPs and comparison of two eye color prediction systems in diverse Eurasia populations. *International Journal of Legal Medicine, 128*, 447–453.

Zaumsegel, D., Rothschild, M. A., & Schneider, P. M. (2013). A 21 marker insertion deletion polymorphism panel to study biogeographic ancestry. *Forensic Science International Genetics, 7*, 305–312.

Zhao, Z., Fu, Y. X., Hewett-Emmett, D., & Boerwinkle, E. (2003). Investigating single nucleotide polymorphism (SNP) density in the human genome and its implications for molecular evolution. *Gene, 312*, 207–213.

Zhao, G. B., Ma, G. J., Zhang, C., Kang, K. L., Li, S. J., & Wang, L. (2021). BGISEQ-500RS sequencing of a 448-plex SNP panel for forensic individual identification and kinship analysis. *Forensic Science International Genetics, 55*, 102580.

Zhao, S., Shi, C. M., Ma, L., Liu, Q., Liu, Y., Wu, F., Chi, L., & Chen, H. (2019). AIM-SNPtag: A computationally efficient approach for developing ancestry-informative SNP panels. *Forensic Science International Genetics, 38*, 245–253.

Zheng, W., He, G., Tao, L., Zhao, X., Liu, J., Wang, M., Zhou, D., Chen, X., Li, C., & Hou, Y. (2018). Massively parallel sequencing of 165 ancestry informative SNPs in two Chinese Tibetan–Burmese minority ethnicities. *Forensic Science International Genetics, 34*, 141.

Zidkova, A., Horinek, A., Stenzl, V., & Korabecna, M. (2013). Application of multifactor dimensionality reduction analysis and Bayesian networks for eye color and ancestry prediction for forensic purposes in the Czech Republic. *Forensic Science International Genetics, 4*.

This page intentionally left blank

CHAPTER 14

Forensic analysis of hair using NGS technology

Ruth M. Benitez and Kelly M. Elkins
Department of Chemistry, Forensic Science Program, Towson University, Towson, MD, United States

Introduction to hair analysis in forensic science

Humans shed a normal amount of 50–100 hairs per day (Aparecida de Franca et al., 2015). Hair samples are a common type of trace evidence encountered at crime scenes. Hair samples that are collected at crime scenes represent about 90% of the evidence most commonly found (Amory et al., 2007). Hair evidence plays an important part in connecting the suspect to the victim or vice versa. It can help investigators with leads of who was, and who was not, present at the crime scene.

DNA analysis is a popular method of identification and can be performed on hair evidence if the item can first be identified as human hair. Hair evidence samples are examined microscopically by forensic examiners to differentiate between human and animal hair. Human and animal hairs are composed of three basic regions: the cuticle, medulla, and cortex (Deedrick & Koch, 2004). These regions vary and have different morphological characteristics among various animal species and humans. Other morphological characteristics such as color, pigmentation, and the overall diameter of the hair shaft can help in differentiating animal from human hair (Deedrick & Koch, 2004). Microscopy examination of hairs can sometimes be viewed as subjective with very minimal statistical analysis (Manheim, 2015). A study conducted in 2015 used attenuated total reflectance Fourier transform infrared (ATR FTIR) spectroscopy to determine a more preliminary objective approach to identifying human hair from animal hairs and fibers (Manheim, 2015). The study performed hair analysis on human, dog, and cat hairs by collecting 10 spectra for each individual and animal. The tool was concluded and demonstrated to be effective in identifying human hair from dog and cat hair (Manheim, 2015). ATR-FTIR can be a good method for aiding in the identification of human hair due to its quick and nondestructive nature toward forensic evidence. This nondestructive technique is important when human hair evidence is sent for DNA analysis for further identification of an individual. Having a well-preserved hair sample improves the chances of obtaining a good DNA profile for an individual.

Next Generation Sequencing (NGS) Technology in DNA Analysis
ISBN 978-0-323-99144-5, https://doi.org/10.1016/B978-0-323-99144-5.00021-4
© 2024 Elsevier Inc.
All rights reserved.

Mitochondrial DNA analysis

Mitochondrial DNA (mtDNA) extraction has been a conventional technique in DNA extraction from hairs when a root is not present for nuclear DNA (nuDNA) extraction. Due to the high copy number of the mitochondrial organelle, mtDNA is the most abundant DNA along a hair shaft, and the shaft is the most frequently obtained region from which DNA can be extracted for analysis. Hair contains cortical cells, which are a source of mtDNA along the hair shaft (Roberts & Calloway, 2007). These cells originate from the germinal bulb matrix cells and from melanocytes (Roberts & Calloway, 2007). Mitochondria are found in the keratinized hair shaft (Linch et al., 2001). Multiple copies of mtDNA from the mitochondria are present along the hair shaft, making it an abundant source of mtDNA (Linch et al., 2001).

mtDNA analysis has been efficient and effective, but compared to nuDNA, which has a higher discriminatory power, mtDNA variation is more limited and is largely limited to tracing the maternal lineage of an individual. An individual inherits a single mtDNA haplotype from their mother (Linch et al., 2001). Some individuals are found to have heteroplasmy at a single locus or throughout the profile. Heteroplasmy can be defined as a heterogenous pool of mtDNA in the cell (Linch et al., 2001). Typically, the control mtDNA region is profiled, and two or more mtDNA populations are observed in a single individual in cases of heteroplasmy (Hühne et al., 1999).

Though mtDNA lacks the discrimination power of nuDNA, it can be used to obtain DNA profiles to include/exclude individuals. The Scientific Working Group on DNA Analysis Methods (SWGDAM) has stated that the discriminatory power of mtDNA is dependent on the reported sequence range of highly variable (HV) 1 and HV2 regions (Damaso et al., 2021). The longer the length sequence, the more likely a variation will be detected. Though these two regions have a lower power of discrimination than short tandem repeats (STRs), they still allow for inclusion or exclusion based on haplotypes (Damaso et al., 2021).

Hair can be recovered in one of three phases: anagen (growth), catagen (preparing to be shed), and telogen (shedding). mtDNA has been successfully recovered from telogen hairs in 92.5% of cases (Brandhagen et al., 2018). The quantity and quality of mtDNA decrease substantially along the length of the hair shaft (Brandhagen et al., 2018). This is seen as the most challenging when recovering mtDNA from old hairs rather than recently shed hairs. The small quantities of mtDNA recovered from old hairs present difficulties in obtaining a probative mtDNA profile in forensic cases involving older, damaged, or degraded hairs compared to recently shed hairs (Brandhagen et al., 2018). Additional factors that contribute to the successful amplification and typing of mtDNA are the donor's race and age due to pigmentation differences (Roberts & Calloway, 2007). Studies have demonstrated that using shotgun sequencing can become an alternative efficient method to successfully obtain complete mtDNA genomes from smaller fragments of very old hairs without the presence of a root (Brandhagen et al., 2018).

Studies have also been conducted on the effects of chemically treating hair and DNA analysis. In 2017, one research study explored the effects of hair-color dyes or treatments

and the impacts the dyes can have on DNA analysis (Al-Sammarraie et al., 2017). In the study, results indicated that hair dye and henna-treated hair impacted DNA analysis results (Al-Sammarraie et al., 2017). The study indicated that henna led to a higher recovered DNA concentration as compared to chemically treated hair (Al-Sammarraie et al., 2017). One explanation may be that the chemicals in the hair dyes, such as hydrogen peroxide, attack the bonds within DNA bases and fragment the DNA (Aparecida de Franca et al., 2015). Chemical treatment can also expose hair to harsh conditions such as UV light, heat, and other chemicals that are known to degrade DNA (Damaso et al., 2021). When scientists attempted secondary and third reamplification and repurification of the color-treated hairs, they were finally able to obtain satisfactory mtDNA results (Al-Sammarraie et al., 2017). Melton et al. (2005) stated that having a history of extreme environmental exposures in hair results in no mtDNA profile. Extracting mtDNA from degraded, old, and chemically treated hairs is more challenging than successfully extracting mtDNA from recently shed hairs. In challenging circumstances, using a method to target smaller mtDNA sequences and using mini primer sets along with the use of next-generation sequencing (NGS) will lead to a better collection of mtDNA sequence information (Damaso et al., 2021).

Nuclear DNA analysis

If a hair root is present, nuDNA extraction is performed. Unfortunately, this is a challenge as 95% of the hairs found at the crime scene are telogen hairs without roots (Lepez et al., 2014). Finding hair with roots poses a challenge when most hairs found at crime scenes are rootless. A 2014 study aimed at developing a fast-screening method for identifying hair roots for successful nuDNA analysis (Lepez et al., 2014). 4′,6-diamidino-2-phenylindole (DAPI) was used to stain nuDNA in hair roots collected from a crime scene (Lepez et al., 2014). Staining can be used to help predict the success of STR analysis (Lepez et al., 2014). The method was demonstrated to be nondestructive, quick, and inexpensive allowing forensic laboratories to analyze only the most promising hair roots containing nuclei (Lepez et al., 2014).

As the majority of crime scene hairs are rootless, recent research has focused on obtaining nuDNA profiles from hair shafts and has led to the success of nuDNA extraction from rootless hairs (Brandhagen et al., 2018; Grisedale et al., 2018; Jehaes et al., 1998). As previously described, hairs are made of dead skin cells that undergo a keratinization process (Aparecida de Franca et al., 2015). Due to the keratinization process, in which cellular organelles and nucleic acids are degraded, it is difficult to obtain nuDNA from hair shafts (Brandhagen et al., 2018). The process can lead to DNA degradation, particularly nuDNA destruction, within the hair shaft (Brandhagen et al., 2018). Although nuDNA has also been obtained from telogen hairs, it is present in low quantities and varies substantially in quality (Brandhagen et al., 2018). Thus, it is not common practice for forensic crime labs to perform nuDNA extraction on telogen hairs simply because qPCR assays in implementation are not sensitive enough to detect the low

quantity levels of nuDNA (Brandhagen et al., 2018). A study did demonstrate that direct amplification of telogen hairs combined with elevated PCR cycles increases the amplicons to lead to full nuDNA STR profiles approximately 20% of the time (Grisedale et al., 2018). The nuDNA typing assay, InnoTyper 21, has been tested and has been shown to improve the success rate for DNA typing of degraded rootless hair shafts (Grisedale et al., 2018). InnoTyper 21 is a small amplicon (60−125 bp) test kit that is reported to be very sensitive and provides useful nuDNA profiles from as little as 50 pg of degraded DNA (Grisedale et al., 2018). To determine the success of the InnoTyper 21 kit, a subgroup of hair samples was tested with the kit, and another was amplified using the GlobalFiler PCR amplification kit following the manufacturer-recommended protocol (Grisedale et al., 2018). The results showed that InnoTyper 21 was more successful in obtaining greater allele recoveries than Globalfiler (Grisedale et al., 2018). Overall, 42% of all hair samples tested produced genotype recoveries greater than 40% with InnoTyper 21 (Grisedale et al., 2018). Stated differently, 40% of the hair shaft samples led to interpretable DNA profiles using the InnoTyper 21 kit (Grisedale et al., 2018). It was noted that some donors yielded sufficient DNA profiles for all three of their samples, but no DNA was recovered from other donors (Grisedale et al., 2018). All the hairs that were analyzed were the same length and collected in the same manner. This indicates that the variation in nuDNA recovery and typing success is different among individuals and is dependent on the donor of the hair (Grisedale et al., 2018). The success of obtaining results from hair shafts largely depends on the optimization of the extraction technique being used and the fidelity of the sample (Grisedale et al., 2018). Obtaining an interpretable nuDNA profile from a single hair shaft is largely dependent upon the hair donor, the DNA quantity recovered, and the degradation index (DI) (Grisedale et al., 2018). The DI is determined by taking the ratio of the short quantity value over the long quantity value (DI= (short)/(long)) (Grisedale et al., 2018). Highly discriminatory nuDNA was shown to be recovered from rootless hair shafts in the study (Grisedale et al., 2018).

NuDNA has also been "assumed" to be contained in the nuclei of epithelial cells on the outer surface of the hair shaft rather than inside the hair shaft in the cortical cells (Brandhagen et al., 2018). Two studies have challenged these hypotheses and been able to demonstrate success in extracting nuDNA from old and ancient hair samples. NuDNA was successfully extracted from degraded and old hair samples from Siberian mummies as well as historical hair samples from the Romanov family (Amory et al., 2007; Loreille et al., 2022). Both of these studies demonstrated the importance of having effective, efficient, and clean sampling and DNA extraction methods to successfully recover nuDNA from hair shafts. Furthermore, Grisedale et al. (2018) showed high discrimination of nuDNA results for over 40% of the hair shaft samples tested and showed that out of 60 hair samples, 25 generated successful, interpretable DNA profiles.

Application of NGS for hair analysis

Next generation sequencing (NGS) is an emerging technology in forensic science. The impact of NGS in the forensic science community is demonstrating promise in aiding improved DNA hair analysis for both mtDNA and nuDNA. The application of NGS for mtDNA, STR, and single nucleotide polymorphism (SNP) typing has demonstrated great promise for the analysis of forensic evidence, especially in cases in which it is degraded and limited (Shih et al., 2018).

Successful hair nuDNA extractions and case studies have opened the field to future studies and inspired scientists to work to improve the methods and techniques to obtain the best possible yields of nuDNA from hair shafts. Quantities of nuDNA recovered from recent hairs were so low that standard qPCR quantification methods did not yield quant values, but nuDNA quantity was assessed using the ratio of nuclear to mtDNA sequence reads in the data (Brandhagen et al., 2018). The focus of a 2018 study was to examine the distribution of nuDNA along a hair shaft (Brandhagen et al., 2018). The total percentage of sequence reads mapped across the human nuclear genome from the recently collected hairs ranged between 88.4% and 99.5% of the total DNA reads (Brandhagen et al., 2018) The authors reported that 99% of the shotgun reads that were mapped to the human genome were nuDNA sequences (Brandhagen et al., 2018) Surprisingly, these high percentages of nuDNA were observed regardless of the purification protocol (Brandhagen et al., 2018). The average size of nuDNA reads varied between 49 and 88 bp for five to six extractions (Brandhagen et al., 2018). Even in samples in which the nuDNA was determined to be more degraded than the mtDNA, it was found in far higher quantities in any given hair or hair sample (Brandhagen et al., 2018). Since the size of the nuclear genome is much larger than the mtDNA chromosome, even with the large percentages of nuDNA reads, obtaining any reasonable depth of coverage across the genome failed (Brandhagen et al., 2018). An Illumina MiSeq FGx NGS instrument was used, which is a comparatively low throughput/output instrument; a higher throughput and deeper sequencer such as an Illumina NextSeq or HiSeq would lead to more mtDNA and nuDNA data (Brandhagen et al., 2018).

Full STR profiles from telogen hairs present a common challenge due to the integrity of the DNA. In 2019, an NGS-based method termed the maSTR assay was developed to genotype single telogen hairs (Heß et al., 2019). The comparison of capillary electrophoresis (CE)-based STR methods with NGS-based STR methods was conducted to determine the best method for achieving successful STR profiles (Heß et al., 2019). The study concluded that the maSTR assay yielded higher allele recovery than CE-based STR analysis using DNA of single telogen hairs; full or partial profiles were obtained with few allelic dropouts and drop-ins (Heß et al., 2019). Extremely degraded DNA samples were not able to be typed using STRs by CE or NGS (Heß et al., 2019). CE-based profiles resulted in more allelic dropouts and fewer drop-ins than NGS-based profiles

(Heß et al., 2019). A CE-based profile resulted in a false exclusion with only one hair sample due to the very little DNA and strong degradation, making it impossible to genotype (Heß et al., 2019). Overall, this study was successfully able to be the first to calculate likelihood ratios of NGS-based STR profiles of hairs with probabilistic methods in a fully continuous manner using the GenoProof Mixture 3 software (Heß et al., 2019). The correct assignment of DNA profiles from single telogen hairs to their donor profiles is dependent on the quality and integrity of the recovered DNA (Heß et al., 2019). When sufficient quality and quantity of DNA are recovered, the hairs can be assigned to their donor with statistical significance using both CE and NGS.

NGS methods have led to advancements in mtDNA profiling of degraded samples, mixed samples, and hair samples. In contrast to using Sanger Sequencing, NGS can rapidly sequence the whole mitochondrial chromosome of several samples in parallel and produce millions of sequence reads in a single sequencing run (Shih et al., 2018). NGS technologies can generate 25 million−3 billion reads per sequencing run with an output of approximately 15−600 Gb file size (Shih et al., 2018). This large quantity of data aids in providing sufficient sequencing information for robust SNP and STR variant calling (Shih et al., 2018). Being able to analyze the whole mitochondrial genome rather than solely the hypervariable polymorphic regions HVI and HVII can increase the discrimination power and alleviate the problem of common sequences found in the HVI and HVII regions shared by individuals of the same area (Shih et al., 2018). HVI and HVII can often be successfully used in analyzing limited or degraded DNA samples, but NGS techniques are more sensitive than previous methods.

In 2015, a sensitive study was conducted to determine the minimum starting quantity of DNA for the Roche 454 HVI/HVII NGS assay (Kim et al., 2015). The assay was demonstrated to be highly sensitive for sequencing limited DNA quantities ($\sim$ 100 mtDNA copies) (Kim et al., 2015). The assay sensitivity was determined to be 0.5 pg or $\sim$ 500 copies (Kim et al., 2015). The DNA quantity could be potentially decreased to 0.1 or 0.005 pg by eliminating the dilution after the amplification of low-copy-number samples (Heß et al., 2019). Results were obtained from lower amounts of DNA than the recommended amount by the manufacturer (Kim et al., 2015). The assay used in the study was demonstrated to be specific to the targeted HVI/HVII regions and did not produce fragment reads outside these regions (Kim et al., 2015). Sequencing the entire mtDNA genome instead of only HVI/HVII regions adds additional and valuable discriminating power (Kim et al., 2015).

Furthermore, NGS can be used to simultaneously analyze mtDNA polymorphisms, STRs, SNPs, and indels (Shih et al., 2018). The sensitivity of NGS can be used to generate profiles when the sample is below the detection limit for CE (Shih et al., 2018). The analysis of mtDNA polymorphisms with the addition of SNPs has proven

to be valuable when nuDNA is too degraded or not recovered from telogen hair roots and hair shafts (Shih et al., 2018). Custom probe capture panels for targeting the whole mt genome and 426 nuclear SNPs on an Illumina MiSeq platform from a single shotgun library were applied to degraded DNA samples, telogen hair roots, and telogen hair shafts (Shih et al., 2018). Using the DNA probes greatly improved, and allowed for the capture of both mtDNA and nuclear SNP markers from a single shotgun DNA library without the consumption of additional DNA extracts for limited biological forensic samples (Shih et al., 2018). The hybrid probe capture NGS system led to mtDNA sequencing results for telogen hairs of 100% coverage of the mt genome at >100x read depth with $>92\%$ aligned to the revised Cambridge Reference Sequence (Shih et al., 2018). Even when SNP loci dropout occurred in DNA tested from telogen hairs, the partial SNP profile combined with the whole mt genome sequence increased the discriminatory power, which likely approached or exceeded the discriminatory power of 13 STR loci profiles based on the number of SNPs typed (Shih et al., 2018).

SNPs can be typed using NGS to predict externally visible characteristics of an individual from biological samples such as hair collected in criminal investigations and missing person cases (Pośpiech et al., 2022). They can also be used to exclude suspects. Prediction models for hair color, hair shape, hair loss, and hair graying have been developed from genome-wide association studies (GWAS) (Pośpiech et al., 2022). Studies explored the genetic determinants of hair thickness and density traits and associated them with two genes, EDAR and FGFR2; these have been reported to be associated with thick hair in Asian populations (Pośpiech et al., 2022). Two SNP loci found in EDAR (rs3827760 and rs365060) linked to hair thickness in Asians were also associated with head hair density in a Polish study with male participants (Pośpiech et al., 2022). In addition, two SNPs were significant in association with hair thickness (rs731236) and hairiness (rs2238136) (Pośpiech et al., 2022). NGS can augment anthropology methodology in investigations (Pośpiech et al., 2022). Accuracy predictions of highly polygenic human traits can be challenging, requiring a large number of typed SNPs or reducing strict significance criteria (Pośpiech et al., 2022). In a 2016 study, hair features were predicted using GWAS data generated for Latin Americans; the highest accuracy was for hair color and beard thickness, with approximately 40% of the phenotypic variation (Adhikari et al., 2016). Further research with more populations is needed to develop more advanced statistical methods and predictive models for hair thickness and hair density traits (Pośpiech et al., 2022). Identification of significant SNPs is more effective when correlations between DNA variants and the cumulative effects of SNPs are considered (Pośpiech et al., 2022). Features including hair color, thickness, and shape can be predicted using commercial NGS forensic DNA typing kits including the Verogen ForenSeq Signature Prep kit and Imagen kit (Elkins et al., 2023; Jäger et al., 2017).

NGS has been used to develop hybrid methods to generate nuclear genotype profiles from rootless hairs that could identify a suspect or victim (Lewis et al., 2022). A custom

collection of oligonucleotide probes targeting 215 SNP loci for ancestry and identification was prepared (Lewis et al., 2022). The DNA recovery quantity per length of hair varied between individuals and between hairs of the same individual (Lewis et al., 2022). The recovered DNA/cm hair was approximately 20 pg, and the SNPs ranged from 21 to 244 reads per sample (Lewis et al., 2022). Using techniques originally developed for ancient DNA typing, the SNP DNA profiles led to a high level of discrimination between individuals (Lewis et al., 2022). The key to success was using sufficient starting material or more than 2 cm of segments (Lewis et al., 2022). The substantial variation between individuals in hair DNA quality and quantity presented a challenge in generating useful genotype profiles for all the hair donors in the study (Lewis et al., 2022).

Continued research in DNA extraction and amplification methods and the development of NGS tools will aid in DNA hair analysis. The impact of NGS on forensic evidence samples encouraged the United States Federal Bureau of Investigation (FBI) to conduct a validation study on the application of NGS for mtDNA casework within the FBI laboratory (Brandhagen et al., 2020). The validation studies were both developmentally and internally validated and were performed in accordance with the *SWDAM Validation Guidelines for Forensic DNA Analysis Methods* and the FBI's *Quality Assurance Standards for Forensic DNA Testing Laboratories* (Brandhagen et al., 2020). The FBI's work demonstrated the efficiency, accuracy, reproducibility, sensitivity, and reliability of the PowerSeq CRM Nested System chemistry (Brandhagen et al., 2020). The PowerSeq assay demonstrated substantial advantages when it came to laboratory workflow (Brandhagen et al., 2020). The PowerSeq assay can be applied uniformly to all sample types without any changes to workflow or level of effort (Brandhagen et al., 2020). This assay simplified and improved the efficiency of the mtDNA analysis process (Brandhagen et al., 2020). The success of validation was submitted to the National DNA Index System (NDIS) and was accepted and approved in May 2019 (Brandhagen et al., 2020). The successful validation led to the approval of mtDNA control region data generated using the PowerSeq assay to be uploaded to the US Combined DNA Index System (CODIS) (Brandhagen et al., 2020). Sequencing only the control region is cheaper and compatible with traditional HVI/HVII mtDNA typing without adding the complexities that typing the complete mt genome data presents (Brandhagen et al., 2020).

Conclusion

Interest in the genetic analysis of hair is increasing with the more widespread availability of NGS tools. Extraction methods have demonstrated the ability to isolate mtDNA and nuDNA from recent and ancient samples. As NGS continues to improve and advance, future NGS implementation appears promising for augmenting the analysis of DNA for forensic hair evidence samples. As NGS methods continue to be approved for CODIS upload, adoption is expected to increase.

References

Adhikari, K., Fontanil, T., Cal, S., Mendoza-Revilla, J., Fuentes-Guajardo, M., Chacón-Duque, J.-C., Al-Saadi, F., Johansson, J. A., Quinto-Sanchez, M., Acuña-Alonzo, V., Jaramillo, C., Arias, W., Lozano, R. B., Pérez, G. M., Gómez-Valdés, J., Villamil-Ramírez, H., Hunemeier, T., Ramallo, V., & Silva de Cerqueira, C. C. (2016). A genome-wide association scan in admixed Latin Americans identifies loci influencing facial and scalp features. *Nature Communications, 7*, 10815. https://doi.org/10.1038/ncomms10815

Al-Sammarraie, H. K. I., AL-Zubaidi, M. M., Namaa, D. S., Hameed, S. N., Saleh, T. Y., Al-Rubai, H. K., Zageer, D. S., & Bader, A. H. (2017). Effects of hair dyes on sequence analysis of hair mitochondrial DNA hypervariable region 1. *Journal of Biotechnology Research Center, 11*(1), 61−66.

Amory, S., Keyser, C., Crubezy, E., & Ludes, B. (2007). STR typing of ancient DNA extracted from hair shafts of Siberian mummies. *Forensic Science International, 166*(2-3), 218−229. https://doi.org/10.1016/j.forsciint.2006.05.042

Aparecida de Franca, S., Darrio, M. F., Esteves, V. B., Baby, A. R., & Velasco, M. V. R. (2015). Types of hair dye and their mechanisms of action. *Cosmetics, 2*(2), 110−126.

Brandhagen, M. D., Just, R. S., & Irwin, J. A. (2020). Validation of NGS for mitochondrial DNA casework at the FBI laboratory. *Forensic Science International: Genetics, 44*, 102151. https://doi.org/10.1016/j.fsigen.2019.102151

Brandhagen, M. D., Loreille, O., & Irwin, J. A. (2018). Fragmented nuclear DNA is the predominant genetic material in human hair shafts. *Genes, 9*(12), 640. https://doi.org/10.3390/genes9120640

Damaso, N., Ashe, E. C., Meiklejohn, K. A., Kavlick, M. F., & Robertson, J. M. (2021). Comparison of polymerases used for amplification of mitochondrial DNA from challenging hairs and hairs of various treatments. *Forensic Science International: Genetics, 52*, 102484. https://doi.org/10.1016/j.fsigen.2021.102484

Deedrick, D. W., & Koch, S. L. (2004). Microscopy of Hair Part 1: A practical guide and manual for human hairs. *Forensic Science Communications, 6*(1), 218681. https://archives.fbi.gov/archives/about-us/lab/forensic-science-communications/fsc/jan2004/research/2004_01_research01b.htm.

Elkins, K. M., Garloff, A. T., & Zeller, C. B. (2023). Additional predictions for forensic DNA phenotyping of externally visible characteristics using the ForenSeq and Imagen kits. *Journal of Forensic Sciences, 68*(2), 608−613. https://doi.org/10.1111/1556-4029.15215

Grisedale, K. S., Murphy, G. M., Brown, H., Wilson, M. R., & Sinha, S. K. (2018). Successful nuclear DNA profiling of rootless hair shafts: A novel approach. *International Journal of Legal Medicine, 132*(1), 107−115. https://doi.org/10.1007/s00414-017-1698-z

Heß, S. A., Biermann, J. P., Grabmüller, M., Madea, B., Thiele, R., & Jäger, R. (2019). Evaluation of STR profile of single telogen hairs using probabilistic methods. *Forensic Science International: Genetics Supplement Series, 7*(1), 454−456. https://doi.org/10.1016/j.fsigss.2019.10.049

Hühne, J., Pfeiffer, H., Waterkamp, K., & Brinkmann, B. (1999). Mitochondrial DNA in human hair shafts-existence of intra-individual differences? *International Journal of Legal Medicine, 112*(3), 172−175. https://doi.org/10.1007/s004140050226

Jäger, A. C., Alvarez, M. L., Davis, C. P., Guzmán, E., Han, Y., Way, L., Walichiewicz, P., Silva, D., Pham, N., Caves, G., Bruand, J., Schlesinger, F., Pond, S. J. K., Valero, J., Stephens, K. M., & Holt, C. (2017). Developmental validation of the MiSeq FGx Forensic Genomics System for targeted next generation sequencing in forensic DNA casework and database laboratories. *Forensic Science International, 28*, 52−70. https://doi.org/10.1016/j.fsigen.2017.01.011

Jehaes, E., Gilissen, A., Cassiman, J. J., & Decorte, R. (1998). Evaluation of a decontamination protocol for hair shafts before mtDNA sequencing. *Forensic Science International, 94*(1-2), 65−71. https://doi.org/10.1016/s0379-0738(98)00052-8

Kim, H., Erlich, H. A., & Calloway, C. D. (2015). Analysis of mixtures using next generation sequencing of mitochondrial DNA hypervariable regions. *Croatian Medical Journal, 56*(3), 208. https://doi.org/10.3325/cmj.2015.56.208

Lepez, T., Vandewoestyne, M., Van Hoofstat, D., & Deforce, D. (2014). Fast nuclear staining of head hair roots as a screening method for successful STR analysis in forensics. *Forensic Science International: Genetics, 13*, 191–194. https://doi.org/10.1016/j.fsigen.2014.08.003

Lewis, E. J., Benefiel, O., Fronda, M. V., & Podini, D. (2022). Nuclear DNA SNP profiles derived from human hair shaft. *Forensic Science International: Genetics Supplement Series, 8*(11), 333–335. https://doi.org/10.1016/j.fsigss.2022.10.080

Linch, C. A., Whiting, D. A., & Holland, M. M. (2001). Human hair histogenesis for the mitochondrial DNA forensic scientist. *Journal of Forensic Sciences, 46*(4), 844–853.

Loreille, O., Tillmar, A., Brandhagen, M. D., Otterslatter, L., & Irwin, J. A. (2022). Improved DNA extraction and illumia sequencing of DNA recovered from aged rootless hair shafts found in relics associated with the Romanov family. *Genes(Basel), 13*(2), 202. https://doi.org/10.3390/genes13020202

Manheim, J. (2015). *Differentiation of human, animal, and synthetic hair by ATR FTIR spectroscopy*. New York: University at Albany SUNY.

Melton, T., Dimick, G., Higgins, B., Lindstrom, L., & Nelson, K. (2005). Forensic mitochondrial DNA analysis of 691 casework hairs. *Journal of Forensic Sciences, 50*(1), 73–80.

Pośpiech, E., Karłowska-Pik, J., Kukla-Bartoszek, M., Woźniak, A., Boroń, M., Zubańska, M., Jarosz, A., Bronikowska, A., Grzybowski, T., Płoski, R., Spólnicka, M., & Branicki, W. (2022). Overlapping association signals in the genetics of hair-related phenotypes in humans and their relevance to predictive DNA analysis. *Forensic Science International: Genetics, 59*, 102693. https://doi.org/10.1016/j.fsigen.2022.102693

Roberts, K. A., & Calloway, C. (2007). Mitochondrial DNA amplification success rate as a function of hair morphology. *Journal of Forensic Sciences, 52*(1), 40–47. https://doi.org/10.1111/j.1556-4029.2006.00322.x

Shih, S. Y., Bose, N., Gonçalves, A. B. R., Erlich, H. A., & Calloway, C. D. (2018). Applications of probe capture enrichment next generation sequencing for whole mitochondrial genome and 426 nuclear SNPs for forensically challenging samples. *Genes (Basel), 9*(1), 49. https://doi.org/10.3390/genes9010049

CHAPTER 15

Forensic DNA phenotyping using next-generation sequencing

Mahdi Haidar[1], Fatema Mousawi[1] and Abdullah K. Al-Matrouk[2]
[1]Kuwait Identification DNA Laboratory (KIDL), General Department of Criminal Evidence, Ministry of Interior, Kuwait City, Kuwait; [2]Forensic Laboratory Department, General Department of Criminal Evidence, Ministry of Interior, Kuwait City, Kuwait

Introduction

The concept of DNA fingerprinting emerged in 1985 when Professor Sir Alec Jeffreys used a variable number of tandem repeats (VNTRs) to generate individual-specific DNA patterns—known as DNA fingerprints—that could be applied to both human identification and paternity testing (Jeffreys, Wilson et al., 1985). The VNTR technique was first described as an effective tool when it was used to help in immigration cases (Jeffreys, Brookfield et al., 1985) and to solve murder cases (Cobain, 2016). However, studies have highlighted a major limitation of the VNTR technique, which requires high-molecular-weight DNA. Therefore, DNA fragment analysis by the VNTR technique is unsuitable for low-molecular-weight and degraded DNA. Poor DNA quantity and quality are two distinct phenomena that are commonly observed in forensic DNA analysis.

Since the early 1990s, with the invention of the polymerase chain reaction (PCR), DNA analysis has taken an enormous step forward in its ability to generate genetic profiles from small amounts of DNA. The PCR makes it possible to analyze the low amounts of DNA often retrieved from forensic samples. In addition, using short tandem repeat (STR) polymorphism and PCR has contributed to increased discrimination power between forensic DNA samples. Amplification of STRs has facilitated their application in analyzing short DNA fragments, such as degraded DNA. Consequently, STR analysis has replaced the VNTR method to become the standard technique used in the forensic laboratory workflow of human DNA identification (Edwards et al., 1991, 1992). The diversity and high discriminatory power of STR markers allow them to be used not only for human identification but also in population genetics studies and paternity cases. Despite their robustness, using STR markers has several limitations that make it challenging to obtain interpretable STR profiles. These limitations include their large amplicon size, high mutation rates [10^{-3} per locus per generation (Brinkmann et al., 1998)], and artifacts such as stutter peaks. Therefore, single nucleotide polymorphism (SNP) typing is an alternative approach to address these shortcomings of STRs.

Next Generation Sequencing (NGS) Technology in DNA Analysis
ISBN 978-0-323-99144-5, https://doi.org/10.1016/B978-0-323-99144-5.00014-7
© 2024 Elsevier Inc.
All rights reserved.

Compared to STRs, SNPs possess lower mutation rates (10^{-8} per locus per generation) (Durrett & Limic, 2001; Reich et al., 2002; Weber & Wong, 1993).

Moreover, the small size of their PCR amplicons makes SNPs more useful for analyzing highly degraded and low-amount DNA samples than STR amplicons (Budowle, 2004), thus overcoming the major limitation of using STRs for highly degraded DNA samples (Sanchez et al., 2006). This advantage can be used in the forensic analysis of mass disasters and missing-person cases where DNA may be severely degraded. SNP markers also avoid one of the interpretational complexities associated with STR typing by generating DNA profiles without stutter peaks. In addition, several SNPs located within DNA coding or regulatory regions can lead to amino acid substitutions that may modify gene function properties, which, in turn, could result in the expression of distinct visible phenotypic traits. Recent evidence suggests that SNPs are the most prevalent type of genetic variation that significantly affects visible phenotypic traits. This provides a valuable opportunity for forensic investigations to potentially identify an unknown individual from a mere DNA trace, especially when a reference sample is unavailable for a DNA match.

The development of technologies that could determine an individual's genotype by scanning large numbers of SNPs has revolutionized the study of phenotypic traits in forensics. Genome-wide association studies (GWAS) were used to identify SNPs associated with a particular phenotype or disease across the genome of large populations or cohorts of individuals by screening hundreds of thousands to millions of SNPs. Some variants showed a significant association with the phenotype of interest and were further validated and subjected to functional analysis to establish their biological significance. However, GWAS has limitations, including the potential for false-positive findings due to multiple testing, and population stratification. In multiple testing, multiple hypotheses are tested simultaneously, leading to an increased risk of observing significant results by chance alone. In population stratification, the subgroups have different allele frequencies due to differences in ancestry (Haidar et al., 2021), which can lead to spurious associations between SNPs and phenotypes if not adequately controlled (Price et al., 2006). Several improvements in the GWAS methodology have been developed to address these limitations, including using more sophisticated statistical algorithms, applying rigorous quality-control measures to eliminate unreliable SNPs, and incorporating genetic ancestry information to control for population stratification (Wray & Goddard, 2010). The GWAS approach has driven forensic geneticists to study SNP markers that are likely to be associated with physical traits to predict an individual's phenotype. Recently, genetic profiling using SNP markers has been carried out using DNA sequencing technology, such as mini-sequencing, Sanger sequencing, and next-generation sequencing (NGS), the latter representing a milestone in human identification in forensic analysis. This chapter is dedicated to exploring the relationship between phenotypic-informative SNPs and externally visible traits. Additionally, it will delve into the use

of NGS technology in testing these SNP-associated phenotypic inferences in DNA forensics. The expansion on these topics sheds light on their significance in forensic research and analysis.

Progression to next-generation sequencing

In the 1970s, two technologies were introduced to 'read' or sequence DNA, including the dideoxy chain termination method engineered by Sanger, known as Sanger sequencing, and the chemical modification method developed by Maxam and Gilbert, called Maxam—Gilbert sequencing. Both methods use radioactive labels to infer the order of nucleotides. Sanger sequencing involves denaturing the target DNA and annealing it to a tagged primer. The DNA polymerase then extends the primer, incorporating chain-terminating nucleotides to create different-sized fragments (Sanger et al., 1977). In contrast, Maxam-Gilbert sequencing utilizes a terminally tagged double-stranded DNA fragment that is digested by a restriction enzyme. Different combinations of chemicals are then used to create base-specific segments of DNA (Maxam & Gilbert, 1977). Both methods determine base order using electrophoresis to separate DNA fragments on polyacrylamide slab gels.

By 1987, several enhancements had been made to the Sanger method, which mainly involved using a fluorescent dye instead of a radioactive label in a single reaction and developing fragment detection using capillary electrophoresis (CE) (Ansorge et al., 1986; Smith et al., 1985). Subsequently, the first generation of automated fluorescence-based sequencing was developed by Applied Biosystems (ABI) using the Sanger technique (Hood et al., 1987; Smith et al., 1986), which was the quantum leap that changed the progress of DNA sequencing technology. These improvements led to substantial growth in sequencing data, from half a million bases in 1982 to around 10 million bases by 1986 (Shendure et al., 2017). The improvement also included an increase in the number of capillaries in the sequencing instruments. The first capillary sequencer, ABI 310, was limited to one capillary, whereas the latest improved model, AB 3730xl, has 96 capillaries. The latter sequencer can deal with larger fragment sizes and has a shorter run time (Gužvić, 2013). Such improvements significantly impacted the sequencing process of the Human Genome Project, the largest and most valuable project based on the Sanger sequencing technique.

However, the Sanger method has major drawbacks, such as its high cost, low sequencing throughput, laborious operating procedures, and limitations in analyzing complex genomes (Fullwood et al., 2009; Metzker, 2010). These limitations have led to the development of new platforms that can determine an individual's genotype by scanning large numbers of SNPs associated with phenotypic traits. The past few decades have witnessed the introduction of massively parallel sequencing, also termed NGS, with an improved DNA sequencing technique that outperforms the traditional Sanger

method in terms of cost, throughput capacity, and sequencing efficiency. NGS is a high-throughput sequencing technology that sequences DNA templates along the human genome in a massively parallel manner to generate millions of DNA sequences (reads). The introduction of NGS has allowed rapid progress in many biological fields, such as clinical genetics (Krier et al., 2016; McCarthy et al., 2013), human history (Marciniak & Perry, 2017), ancient DNA analysis (Poinar et al., 2006), microbiology (Motro & Moran-Gilad, 2017), genetics (Abecasis et al., 2012), and forensic science (Palencia-Madrid & de Pancorbo, 2015; Weber-Lehmann et al., 2014). NGS technologies are routinely used to sequence whole human genomes as well as target regions to permit the sequencing of a subset of the whole genome. Targeted NGS technologies are capable of analyzing large numbers of markers in a single run, including SNPs, STRs, insertion and deletion polymorphisms, structural variants, and copy number variants (Metzker, 2010).

Phenotypic-informative SNP in forensics

Several studies have shown that SNPs are the most common type of genetic variation marker associated with certain physical traits of individuals that can serve as a forensic tool. Due to their low mutation rates, SNPs are inherited in stable patterns across generations, making them suitable markers for studying ancestral relationships and for use in phenotypic prediction studies (Budowle & Van Daal, 2008). SNP markers that are associated with a particular physical trait can be used to predict a person's physical appearance and are commonly referred to as phenotypic-informative SNPs. Hence, this process is known as forensic DNA phenotyping (FDP). It has been reported that FDP using phenotypic-informative SNP markers can act as "biological eyewitnesses," predicting certain crime suspect/perpetrator phenotypes and supporting human witnesses to identify an individual (Kayser, 2015; Valentine & Fitzgerald, 2016). Phenotype inference facilitates the prediction of physical attributes including hair, eye, and skin pigmentation, in addition to facial morphology, all derived from DNA samples. This technology has several potential applications in forensic investigations, including identifying individuals and establishing their physical characteristics when traditional methods have proven ineffective.

As previously discussed, NGS technology significantly facilitates the sequencing of large amounts of DNA spanning an individual's genome. This includes regions that encapsulate data pertaining to the physical attributes of the individual. One of the pivotal methods employed for phenotype inference via NGS involves sequencing targeted phenotypic-informative SNPs associated with specific physical traits. This technique aims to identify an unknown individual by forecasting their physical characteristics derived from a DNA sample. This technology has contributed to narrowing down lists of suspects or victims and expediting investigative timelines. Furthermore, phenotype

inference proves invaluable in the identification of human remains, particularly when traditional identification methods like STR analysis are rendered infeasible due to factors such as decomposition or the unavailability of dental records or fingerprints.

Externally visible characteristics

As highlighted in previous sections, the FDP is defined as predicting an individual's externally visible characteristics (EVCs) through the use of informative genetic SNPs. This technique has been particularly instrumental for forensic applications, aiding in the identification of both suspects and victims in criminal investigations. Nonetheless, it is essential to note that this DNA intelligence approach functions primarily as an investigative tool as opposed to a method for identifying specific individuals (Kayser, 2015). This distinction stems from its predominant use in excluding suspects from a crime scene investigation where their DNA does not correspond with the EVCs inferred from DNA samples procured from the scene. Such technology undoubtedly bolsters law enforcement efforts, particularly in criminal investigations and missing person cases involving unidentified individuals. Indeed, the FDP methodology has proven successful in identifying human remains and victims of disasters. Notable instances of successful FDP applications include the resolution of the Louisiana serial killer case (Fullwiley, 2008) and the Madrid train attacks (Phillips et al., 2007, 2009). Furthermore, the utility of FDP extends beyond humans, finding application in breed classification and skeletal phenotype prediction based on canine DNA evidence (Raymond et al., 2022).

Several studies have shown the association between specific phenotypic traits and their genetic determinants in forensics. However, several challenges necessitate resolution prior to the routine incorporation of FDP into investigative procedures. For example, there are limitations to the accuracy and precision of current predictions of certain FDP tests, which, in turn, are reflected by the average prediction accuracy of EVCs. Research has identified variations in the accuracy of different attributes. These include the predictive accuracy of SNPs associated with factors such as height (Aulchenko et al., 2009; Liu et al., 2014), male pattern baldness (Redler et al., 2012; Richards et al., 2008), hair morphology (Fujimoto et al., 2008, 2009; Medland et al., 2009; Pośpiech et al., 2018), and facial structure (Boehringer et al., 2011; Claes et al., 2014; Sheehan & Nachman, 2014).

Physical traits such as hair and eye color and facial structure are complex because multiple genes control them. They are also influenced by environmental factors such as climate, ultraviolet (UV) light exposure, and nutrition. Consequently, their variation is continuous. This results in a continuum of variation, and, as such, DNA tests cannot account for these nongenetic components. Apart from genetic heritability and environmental effects, several other factors can impact the prediction accuracy of EVCs. These encompass the DNA selection and genotyping method, the informativeness of

the SNPs, and the reference dataset employed. Given these intricacies, it is crucial to address certain legal and ethical considerations when deploying NGS technology for phenotype inference. Such technology must adhere to legal and ethical principles to protect an individual's privacy and dignity. It is essential to avoid discriminatory practices, segregation, biased judgments, or ethnic persecution based on phenotype inference for specific population groups (Schneider et al., 2019). Governments should ensure that NGS-based phenotype inference is conducted in compliance with the due process of law and by forensic professionals who uphold rigorous standards.

More research and collaborative efforts are needed for a comprehensive understanding of predictive SNPs associated with complex EVCs. This enhanced understanding stands to augment the accuracy and precision of FDP. Forensic DNA tests for analyzing DNA markers that cause or are associated with particular features, such as eye, hair, and skin color, have been developed and validated using suitable statistical predictive models (Boehringer et al., 2011; Budowle & Van Daal, 2008). Furthermore, pigmentation traits are among the least genetically complex of all EVCs, with a majority of the phenotypic information being conveyed by a modest number of genes. As a result, the prediction of traits associated with these features yields relatively accurate outcomes (Boehringer et al., 2011; Budowle & Van Daal, 2008). The EVCs discussed in this chapter, including eye, hair, and skin color, along with facial features, represent the most extensively researched in the field of forensics. These EVCs are particularly important as they furnish the most informed determinations in investigations when in pursuit of and identifying suspects.

Eye color

Eye color is an easily observable yet genetically complex characteristic. The study of eye color genetics has a rich history, dating back to the early 1900s, when scientists first began examining the inheritance patterns of eye color (Davenport & Davenport, 1907). The color of an individual's iris is determined by the production and distribution of pigment, which are controlled by multiple genes and factors. Understanding the genetics of eye color has significant implications, including identifying disease risks, developing personalized medicines, and conducting forensic investigations.

The OCA2 gene on Chromosome 15 is the most significant contributor to human eye color variation (Sturm & Larsson, 2009). This gene codes for a protein involved in melanin pigment formation and distribution in the iris. OCA2 SNPs are significantly associated with determining human eye color (Branicki et al., 2008; Duffy et al., 2007). Further research using GWAS and linkage analysis has identified the HERC2 gene, located near OCA2 and specifically linked to blue eyes (Kayser et al., 2008). In addition to blue, various eye colors, such as green, gray, and hazel, are observed in the human population. These colors are thought to result from the interplay of multiple genetic factors and the amount and type of melanin in the iris (Sturm et al., 2008). One

variation in HERC2, rs12913832, is particularly linked to blue eyes, affecting OCA2 expression regulation. Individuals with two copies of the "A" allele of rs12913832 tend to have less melanin in their irizes and are more likely to have blue eyes. This allele is more common in individuals of European descent, which could explain the higher proportion of blue-eyed individuals in that population (Donnelly et al., 2012; Sturm et al., 2008).

IrisPlex, the first highly sensitive multiple genotyping assay for predicting blue/brown eye color, marked a significant milestone in eye color prediction (Walsh et al., 2011, 2012). The IrisPlex system targets six SNPs—rs12913832 (HERC2), rs1800407 (OCA2), rs12896399 (LOC105370627), rs16891982 (SLC45A2), rs1393350 (TYR), and rs12203592 (IRF4)—that are involved in eye color determination in the European population. These SNPs were selected based on their linkage with specific eye colors and allele frequencies in different populations. However, IrisPlex's accuracy in predicting intermediate eye colors is lower than for blue and brown eyes (Liu et al., 2009; Walsh et al., 2012). Furthermore, its efficacy in predicting blue/brown eye color decreases when tested on more heterogeneous samples, such as those with Asian, African, and South American ancestry (Keating et al., 2013). The IrisPlex genetic test, initially developed based on a Dutch database, has shown varying degrees of predictive accuracy across different European subpopulations, with slightly lower effectiveness in individuals from Southern Europe than those from Western Europe (Dario et al., 2015; Purps et al., 2011; Salvoro et al., 2019). Therefore, the identification of the optimal SNPs for determining EVCs is markedly influenced by population-specific factors, indicating that the panels employed in predicting eye color should be tailored to specific geographical regions. Future research directions in the field include refining prediction models, expanding them to include additional populations, and exploring the use of artificial intelligence and machine learning to enhance eye color prediction accuracy.

Hair color

Hair pigmentation is one of the most visible phenotypic traits in human appearance, with significant variation observed between individuals. The hair phenotype can be used in forensic contexts for identification purposes and has been a subject of interest to genetics researchers for years. The pigmentation of human hair involves various biological processes that determine its color, texture, and growth rate. At the core of hair pigmentation is melanin, a pigment produced in the melanocyte cells of the hair follicle. Melanin synthesis occurs via two distinct pathways: Eumelanin, which results in brown and black hair, and pheomelanin, which determines red and yellow hair. The primary gene responsible for determining hair color is the melanocortin 1 receptor gene (MC1R), which regulates melanocyte production of melanin pigments. Studies have shown that variants of MC1R have a significant impact on the expression of these pigments, and therefore on

hair color; for instance, a common MC1R mutation is the R151C gene variant, which decreases the eumelanin pigmentation and increases the pheomelanin pigmentation, resulting in blonde or red hair (Duffy et al., 2004). Interestingly, some genes that determine hair pigmentation are also linked to skin pigmentation, suggesting a relationship between these two characteristics. The SLC24A5 gene, for example, plays a role in both the production of melanin in the skin and melanocytes, and so it is associated with skin pigmentation as well as lighter hair color (Lamason et al., 2005). Forensic scientists target specific SNPs that correlate with hair color. These SNPs include MC1R, TYR, TYRP1, SLC24A5, SLC45A2, OCA2, and HERC2 (Valenzuela et al., 2010; van Daal, 2008). The MC1R SNP is one of the most common and significant SNPs used to determine hair color. Variants of this SNP have been found to influence the expression of pheomelanin and eumelanin (i.e., the pigments responsible for red/yellow and brown/black hair, respectively) (Rees, 2003).

A new forensic phenotyping assay, HIrisPlex, has been developed as an enhanced version of IrisPlex (Walsh et al., 2013). HIrisPlex comprises 24 forensically validated phenotypic markers in a single multiplex, including 18 new hair-color polymorphisms (10 MC1R SNPs, TUBB3, 2 SLC45A2 SNPs, KITLG, LOC105374875, SLC24A4, PIGU, and TYRP1), along with the 6 IrisPlex SNPs (Walsh et al., 2013). These markers have been successfully used several times as predictive tools for analyzing ancient skeletal remains (Chaitanya et al., 2017; Draus-Barini et al., 2013; Haeusler et al., 2016; King et al., 2014). The HIrisPlex prediction has been evaluated for individuals across Europe. The prognostication accuracies are nearly 70% for blond hair, 80% for brown, 80% for red, and 88% for black for individuals from various regions in Europe. In order to establish its validity beyond European populations, Walsh et al. applied the HIrisPlex analysis to DNA samples from the HGDP-CEPH human genome diversity cell line panel, consisting of global datasets representing 51 distinct ethnicities (Walsh et al., 2014). The study found that the strength of HIrisPlex eye and hair color probabilities can enable an inference accuracy exceeding 86% regarding whether a person with brown eyes and black hair originates from nonEuropean regions or European regions. This information holds the potential to enhance forensic phenotyping in the future.

One of the advantages of the HIrisPlex genotyping assay is that its sensitivity exceeds that of numerous other methods used in forensic settings because it can produce complete profiles down to a minimum 63-pg DNA input. In addition, it has been demonstrated to yield precise and trustworthy results when employed on simulated casework specimens composed of diverse biological materials, such as blood, semen, saliva, hair, and trace amounts of DNA (Walsh et al., 2014). This holds promise for use in identifying missing persons or victims of disasters when antemortem samples and possible relatives are not accessible. By using DNA-predicted characteristics, such as eye and hair color, such individuals may be located and making them available for STR profiling that can enable their accurate identification. However, it is essential to consider the limitations of the

HIrisPlex genotyping assay. For instance, while the system has demonstrated high accuracy in predicting hair and eye color, it may be less effective in populations with less-studied genetic backgrounds. Furthermore, the accuracy of the predictions can be influenced by various factors, such as environmental conditions, age, and individual genetic variations. Therefore, future research should continue to refine and expand the scope of the HIrisPlex assay to improve its accuracy and applicability in a broader range of forensic contexts.

Skin color

Skin color is a complex trait determined by various genetic factors and environmental influences. The primary determinant of skin color is the amount and distribution of melanin produced by melanocytes, which are specialized skin cells in the epidermis basal layer. Melanocytes produce melanin in specialized organelles called melanosomes, which are transferred to the surrounding skin cells to protect against UV radiation. The most significant genetic factor affecting skin color is the 1MC1R gene, which regulates the production of eumelanin. Variations in the MC1R gene can result in different levels of eumelanin production, leading to various skin colors (Kanetsky et al., 2004). Moreover, research has shown that variations in the HERC2 gene are associated with light and dark skin colors. Other genes that play a role in skin color include SLC24A5, SLC45A2, and TYR, which can initiate melanin production, and OCA2, which regulates the size and distribution of melanosomes (Andrade et al., 2017; Pośpiech et al., 2014; Soejima & Koda, 2007).

Additional genes, such as KIT ligand (KITLG) and agouti signaling protein (ASIP), also play significant roles in the variation of skin pigmentation. KITLG plays a crucial role in the development and migration of melanocytes, ultimately influencing skin color (Wang et al., 2021). Conversely, ASIP contributes to skin pigmentation by regulating the production of eumelanin and pheomelanin (Bonilla et al., 2005). These two forms of melanin are responsible for the coloration of skin, hair, and eyes. Environmental factors, such as exposure to sunlight and UV radiation, can also influence skin color. UV radiation stimulates melanocytes to produce more melanin, resulting in skin darkening or tanning as a protective measure against DNA damage and skin cancer (Jablonski & Chaplin, 2010). Over time, populations living closer to the equator, where UV radiation is more intense, have developed darker skin, while populations living farther from the equator, where UV radiation is less intense, have developed lighter skin.

A novel tool called HIrisPlex-S, which simultaneously predicts eye, hair, and skin color, was developed as an advancement of the HIrisPlex system. HIrisPlex-S employs 36 SNPs for making skin color predictions, incorporating 19 from the previous model and introducing 17 new markers, including ANKRD11, ASIP, BNC2, DEF8, 4 HERC2 SNPs, MC1R, 4 OCA2 SNPs, RALY, SLC24A4, SLC24A5, and TYR. HIrisPlex-S

has demonstrated its ability to predict skin color on a worldwide scale, achieving high accuracy rates in predicting dark, intermediate, pale, and very pale colors (Breslin et al., 2019).

Face feature

Facial morphology prediction is a process that has gained immense interest in various scientific fields, including genetics, anthropology, and forensics. As a powerful tool with significant potential for its application in forensic contexts and various investigative purposes, it can aid law enforcement agencies in identifying suspects, reconstructing victims' appearances, and facilitating missing person investigations. However, using genetic markers to accurately predict facial traits is much more complex than eye, hair, and skin pigmentation. This complexity is due to the contribution of multiple genes associated with facial feature diversity. Therefore, accurate or reliable prediction of facial morphology continues to remain challenging. However, over recent years, significant advances in genomics have enabled the identification of key genes and genetic variants linked to facial feature development. Some SNPs within these genes have been analyzed for their impact on facial measurements, such as the distance between eyeballs, nose morphology, and mandible structure (Boehringer et al., 2011).

These advances have primarily stemmed from integrating sequencing technologies and GWAS scanning. NGS has allowed for rapid and cost-effective sequencing of large populations, which in turn has facilitated the discovery of genetic variants linked to facial features. Coupled with GWAS data, scientists have developed computational algorithms, statistical models, and high-resolution three-dimensional imaging that greatly enhance the capacity to predict facial landmarks in great detail and with approximate accuracy (Shaffer et al., 2016). Initial breakthroughs in understanding facial morphology were achieved through two GWAS studies conducted on Europeans (Liu et al., 2012; Paternoster et al., 2012). They identified a significant association between the rs7559271 SNP, located within an intron of the PAX3 gene, and nasal bridge morphology. This association was also observed in other populations, highlighting PAX3's significant role in facial development (Adhikari et al., 2016). Furthermore, four novel genetic variants close to the genes PRDM16, TP63, C5orf50, and COL17A1 have been found to influence facial structure development (Liu et al., 2012). These data imply that such genetic variants are essential for face development and involve the normal variation of human facial morphology.

Moreover, a study has investigated the genetics behind facial shape in individuals of European ancestry and identified two novel genetic associations:SNP rs9456748 within an intron of PARK2 with midface height and SNP rs72713618 within FREM1 with central upper lip height (Lee et al., 2017). While the biological role of PARK2 in craniofacial development remains unclear, evidence suggests that FREM1 may influence facial

variation. These findings highlight the potential of data-driven multivariate phenotyping in facial morphology research. In terms of nose shape, a separate GWAS study reported five genetic variants located near the genes DCHS2, RUNX2, GLI3, PAX1, and EDAR (Adhikari et al., 2016). Each of these genes plays a role in various developmental processes that influence the formation and shape of the nose. For example, PAX1 contributes to skeletal development, potentially affecting nasal bone and cartilage, while EDAR affects ectodermal development, impacting the external part of the nose.

There is strong evidence that the development of facial traits undergoes genetic regulation to create an inherited, unique, and recognizable face, which plays an obvious role in physical identification. For example, studying the genetic basis of facial development in African children and adolescents identified two genes, SCHIP1 and PDE8A, associated with facial size, in addition to ten other loci linked to facial shape (Cole et al., 2016). Mouse models further supported these findings, in which genetic variants of SCHIP1 and PDE8A were involved in mouse facial morphogenesis. This study highlighted differences from previous studies on individuals of European ancestry, suggesting a link between facial traits and population genetic structure and providing evidence that ancestry could affect facial morphology. Indeed, the importance of ancestry markers in determining facial features was demonstrated, and the researchers were able to predict facial feature models with a certain degree of reliability by constructing a face using these ancestry markers and then adding other SNPs (Claes et al., 2014, 2018). These models have successfully predicted specific facial features such as eyebrow width and the distance between the eyes.

Additionally, a study on the Korean population identified five significant genetic variants associated with facial morphology. For example, the OSR1-WDR35 locus (rs7567283, G allele) was associated with the frontal facial contour; HOXD1-MTX2 (rs970797, A allele) as well as WDR27 (rs3736712, C allele) loci were associated with eye shape; and SOX9 (rs2193054, C allele) as well as DHX35 (rs2206437, A allele) loci were associated with nose shape, specifically nasal angle and subnasal width, respectively. Comparing with the collective data from the genome-wide project, this study showed that allele frequencies of four SNPs associated with nose shape have differed among ethnic populations (Cha et al., 2018), suggesting that the shape of the noseis genetically inherited.

The Snapshot FDP System is a DNA analysis tool developed by Parabon Nanolabs that can predict facial morphology, sex, eye, hair, and skin color, freckling, and biogeographic ancestry in a single analysis. The system uses DNA samples from crime scenes to profile a potential suspect's appearance. Interestingly, some law enforcement agencies have used the system to solve cases. Nevertheless, further research is needed to identify additional SNPs associated with facial features to improve the system's accuracy level, which may vary depending on the quality and quantity of the DNA sample.

Next-generation sequencing in forensics

The PCR-CE-based analysis methods are widely used in forensic laboratories for typing forensic genetic markers, such as STRs, SNPs, and recently introduced indel markers (Haidar et al., 2020; LaRue et al., 2012). STRs are the most discriminative and commonly used markers in forensics due to their multiallelic state. In contrast, most SNPs exist in a diallelic state, which limits their discriminatory power. It is estimated that 50 SNPs may be as informative as only 13 STRs, meaning that a single STR can match the identification power of 4–5 SNPs (Gill, 2001; Pakstis et al., 2010). Forensic genetic markers, including STRs and SNPs, are found in coding and noncoding regions of the human genome. However, they require individual assays for each class of markers, as they cannot be typed together. This necessitates additional PCR assays and more of the often-limited DNA extract from forensic samples. Moreover, SNP markers have limitations in forensic analysis, such as requiring more DNA than STRs and more labor-intensive analysis methods, like Sanger sequencing, which is impractical for forensic analysts. It also requires a high level of expertize to perform and interpret, which might not always be available in a forensic lab.

NGS technology has advanced forensic practice by allowing for the coamplification and individual sequencing of hundreds of SNPs, in addition to other classes of forensic genetic markers, in a single assay. This saves time, especially when additional DNA investigations are needed. In addition, unlike the CE method, which only measures the length of the overall PCR products and can result in partial profiles, NGS provides additional data measurement by analyzing the targeted STR alleles' sequence and accompanying stutter products. NGS can coamplify hundreds of different types of markers in a single run, marking a new era of high-resolution sequencing in forensic applications. In addition, it enables simultaneous analysis of a combination of STRs and different SNP classes, such as biogeographical ancestry SNPs (associated with a person's geographical origin) and phenotypic SNPs of EVCs, in a single sequencing run. This substantially increases the sequencing data's overall informativeness and forensic testing's discriminatory power while decreasing match probabilities by several orders of magnitude.

One significant advantage of sequencing technologies is the ability to utilize haplotypes, consisting of groups of closely linked SNPs on a single chromosome inherited as a single unit. This approach enables a more comprehensive understanding of genetic information and relationships. In the context of forensic science, microhaplotypes, also known as microhaps, have emerged as valuable tools for human identification. These multi-SNP haplotypes comprise two or more closely linked SNPs separated by less than 200 base pairs (Kidd et al., 2018; Oldoni et al., 2019). Microhaplotypes have gained recent attention from the forensics community as promising alternatives to separately genotyping individual SNPs and STRs for human identification. Microhaplotypes allow forensic scientists to obtain more information from a DNA sample, which leads to

increased accuracy and efficiency in human identification efforts. However, it is essential to note that not all informative-phenotypic SNPs may form phenotypic microhaps. Therefore, accurate phenotype prediction requires genotyping a large number of individual SNPs (Kidd et al., 2014, 2018; Oldoni et al., 2019). Despite this limitation, the ability to analyze both haplotypes and microhaplotypes in conjunction with other genetic markers further demonstrates the versatility and utility of sequencing technologies in forensic applications and potentially generates more information from a smaller amount of starting material compared to some traditional methods.

NGS technologies have become the new standard in the field, with several platforms developed to cater to different sequencing chemistries, data outputs, and applications. Examples of these platforms include Illumina, Ion Torrent, Pacific Biosciences, and Oxford Nanopore Technologies. The choice of the NGS platform depends on factors such as read length, data output, sequencing accuracy, cost, and turnaround time. In forensic applications, NGS has been utilized to simultaneously sequence various markers, such as STRs and SNPs, enabling a more comprehensive analysis of DNA samples. The Ion Torrent platform, developed by Thermo Fisher Scientific, is another widely used NGS technology that relies on a unique sequencing-by-synthesis approach. Unlike the fluorescence-based detection utilized by Illumina, Ion Torrent measures the release of hydrogen ions as nucleotides are incorporated into the growing DNA strand. This method allows for real-time detection of the nucleotide sequence through semiconductor-based sensors. The Ion Torrent platform is known for its rapid sequencing, lower cost, and simplified library preparation compared to other NGS platforms.

For forensic applications, Thermo Fisher Scientific has developed the Precision ID NGS System for human identification, which includes the Ion AmpliSeq DNA panels and the Ion S5 sequencing platform. The Precision ID NGS System is designed to streamline the process of DNA profiling and enable simultaneous analysis of a wide range of markers, including autosomal STRs, Y-STRs, X-STRs, and SNPs. The Ion AmpliSeq DNA panels are a collection of ready-to-use primer sets that target specific regions of interest in forensic genomics. These panels facilitate multiplex PCR amplification of numerous markers in a single reaction, saving time and reducing the required input DNA. The panels available in the Precision ID System include the GlobalFiler NGS STR Panel, the Precision ID Identity Panel, the Precision ID Ancestry Panel, and the Precision ID mtDNA Whole Genome Panel, each designed for specific forensic applications.

The MiSeq FGx forensic genomics system, developed by Illumina and now part of Verogen, is a groundbreaking NGS platform designed for forensic genomic applications. The system leverages the well-established Illumina sequencing technology, which uses a synthesis approach for labeling DNA. This approach involves the incorporation of dNTPs during the DNA polymerization process. As each dNTP is added to the growing

DNA strand, it emits a unique fluorescent signal, which is captured using high-resolution imaging to determine the nucleotide sequence accurately. The MiSeq FGx system offers several advantages: higher throughput, improved sensitivity, and the ability to simultaneously analyze multiple markers, including STRs, SNPs, and other genomic regions of interest. These features make the platform particularly suitable for forensic genomics, where obtaining comprehensive and accurate data from challenging samples is often crucial.

Verogen offers a suite of forensic kits designed to work with the MiSeq FGx system, providing targeted and streamlined solutions for various forensic applications. Essential Verogen forensic kits include the ForenSeq DNA Signature Prep Kit, which enables simultaneous, high-resolution sequencing of 231 markers, comprising 153 identity markers (amelogenin, the sex chromosome marker, 27 autosomal STRs, 24 Y-STRs, 7 X-STRs, and 94 identity SNPs), 22 phenotypic SNPs, and 56 ancestry SNPs. In addition, the ForenSeq Kintelligence Kit targets kinship and intelligence markers, providing information on biogeographical ancestry, EVCs, and genetic relatedness by analyzing 165 markers, including 94 autosomal SNPs, 27 Y-SNPs, 22 X-SNPs, and 22 phenotypic SNPs. Finally, the ForenSeq mtDNA Control Region Kit targets the hypervariable regions of the mitochondrial DNA (mtDNA) control region, offering high-resolution sequencing for human identification and maternal lineage analysis, optimized for challenging samples, and providing robust coverage of the mtDNA control region.

A considerable amount of literature on the ForenSeq kit has been published, assessing the forensic performance of both the kit and the sequencing system, including tests on sensitivity, reproducibility, concordance, and mixture analysis (Almalki et al., 2017; Iozzi et al., 2015; Köcher et al., 2018; Silvia et al., 2017; Xavier & Parson, 2017). Other studies have focused on the applications of the ForenSeq kit in forensic DNA analysis, including its performance on degraded DNA (Calafell et al., 2016; Zhang et al., 2018), kinship testing (Kostrzewa et al., 2017; Ma et al., 2016), and ancestry prediction (King et al., 2018; Ramani et al., 2017). To date, several studies have tested the performance of the ForenSeq markers on different populations (Casals et al., 2017; Churchill et al., 2017; Devesse et al., 2018; Kim et al., 2018; Novroski et al., 2016; Wendt et al., 2016). These studies have demonstrated that sequence variants can be detected in the STRs and their flanking regions (indel or SNP variants). In addition, SNPs can provide additional allelic diversity compared to CE-based methods with their flanking regions. This increased number of genetic variants detected in forensic-marker regions using NGS enhances the power of this kit as a forensic DNA tool.

Forensic laboratories face considerable challenges in implementing NGS technology due to its labor-intensive workflow and the significant expense associated with sequencing, including the cost of equipment and kits. However, as the sequencing cost has dramatically dropped and various NGS systems have become available, the NGS system is expected to become more straightforward and affordable in the coming

years. This shift will undoubtedly accelerate the advancement of forensic genomic applications.

The integration of phenotype prediction with DNA profiling amplifies the investigative potential of forensic genetics. Identifying phenotypic markers through NGS-based SNP analysis has become valuable in forensic science and criminal investigations. While the identification of markers linked to external physical characteristics, such as hair, skin, and eye color, along with facial features, has garnered substantial attention, it is imperative to acknowledge that numerous other SNPs could provide equally beneficial information. Incorporating these additional SNPs into a single NGS kit could provide more comprehensive genetic information for identification purposes, leading to more accurate and efficient forensic investigations and contributing to solving the most challenging criminal cases in the future. A deliberate and cautious approach toward incorporating new phenotypic markers can help maximize the benefits of genetic information in forensic science. Further exploration of the underlying genetic basis of complex physical traits, along with the development and validation of additional markers, may enhance the prediction accuracy of FDP in investigations.

References

Abecasis, G. R. (2012). The 1000 Genomes Project Consortium. An integrated map of genetic variation from 1092 human genomes. *Nature, 491*(7422), 56–65. https://doi.org/10.1038/nature11632

Adhikari, K., Fontanil, T., Cal, S., Mendoza-Revilla, J., Fuentes-Guajardo, M., Chacón-Duque, J. C., Al-Saadi, F., Johansson, J. A., Quinto-Sanchez, M., Acuña-Alonzo, V., Jaramillo, C., Arias, W., Lozano, R. B., Pérez, G. M., Gómez-Valdés, J., Villamil-Ramírez, H., Hunemeier, T., Ramallo, V., De Cerqueira, C. C. S., & Ruiz-Linares, A. (2016). A genome-wide association scan in admixed Latin Americans identifies loci influencing facial and scalp hair features. *Nature Communications,* 7. https://doi.org/10.1038/ncomms10815

Almalki, N., Chow, H. Y., Sharma, V., Hart, K., Siegel, D., & Wurmbach, E. (2017). Systematic assessment of the performance of illumina's MiSeq FGx™ forensic genomics system. *Electrophoresis, 38*(6), 846–854. https://doi.org/10.1002/elps.201600511

Andrade, E. S., Fracasso, N. C. A., Strazza Júnior, P. S., Simões, A. L., & Mendes-Junior, C. T. (2017). Associations of OCA2-HERC2 SNPs and haplotypes with human pigmentation characteristics in the Brazilian population. *Legal Medicine, 24*, 78–83. https://doi.org/10.1016/j.legalmed.2016.12.003

Ansorge, W., Sproat, B. S., Stegemann, J., & Schwager, C. (1986). A non-radioactive automated method for DNA sequence determination. *Journal of Biochemical and Biophysical Methods, 13*(6), 315–323. https://doi.org/10.1016/0165-022X(86)90038-2

Aulchenko, Y. S., Struchalin, M. V., Belonogova, N. M., Axenovich, T. I., Weedon, M. N., Hofman, A., Uitterlinden, A. G., Kayser, M., Oostra, B. A., van Duijn, C. M., Janssens, A. C. J. W., & Borodin, P. M. (2009). Predicting human height by Victorian and genomic methods. *European Journal of Human Genetics, 17*(8), 1070–1075. https://doi.org/10.1038/ejhg.2009.5

Boehringer, S., Van Der Lijn, F., Liu, F., Günther, M., Sinigerova, S., Nowak, S., Ludwig, K. U., Herberz, R., Klein, S., Hofman, A., Uitterlinden, A. G., Niessen, W. J., Breteler, M. M. B., Van Der Lugt, A., Würtz, R. P., Nöthen, M. M., Horsthemke, B., Wieczorek, D., Mangold, E., & Kayser, M. (2011). Genetic determination of human facial morphology: Links between cleft-lips and normal variation. *European Journal of Human Genetics, 19*(11), 1192–1197. https://doi.org/10.1038/ejhg.2011.110

Bonilla, C., Boxill, L.-A., Donald, S. A. M., Williams, T., Sylvester, N., Parra, E. J., Dios, S., Norton, H. L., Shriver, M. D., & Kittles, R. A. (2005). The 8818G allele of the agouti signaling protein (ASIP) gene is ancestral and is associated with darker skin color in African Americans. *Human Genetics, 116*(5), 402–406. https://doi.org/10.1007/s00439-004-1251-2

Branicki, W., Szczerbińska, A., Brudnik, U., Wolańska-Nowak, P., & Kupiec, T. (2008). The OCA2 gene as a marker for eye colour prediction. *Forensic Science International: Genetics Supplement Series, 1*(1), 536–537. https://doi.org/10.1016/j.fsigss.2007.10.062

Breslin, K., Wills, B., Ralf, A., Ventayol Garcia, M., Kukla-Bartoszek, M., Pospiech, E., Freire-Aradas, A., Xavier, C., Ingold, S., de La Puente, M., van der Gaag, K. J., Herrick, N., Haas, C., Parson, W., Phillips, C., Sijen, T., Branicki, W., Walsh, S., & Kayser, M. (2019). HIrisPlex-S system for eye, hair, and skin color prediction from DNA: Massively parallel sequencing solutions for two common forensically used platforms. *Forensic Science International: Genetics, 43*. https://doi.org/10.1016/j.fsigen.2019.102152

Brinkmann, B., Klintschar, M., Neuhuber, F., Hühne, J., & Rolf, B. (1998). Mutation rate in human microsatellites: Influence of the structure and length of the tandem repeat. *The American Journal of Human Genetics, 62*(6), 1408–1415. https://doi.org/10.1086/301869

Budowle, B. (2004). SNP typing strategies. *Forensic Science International, 146*, S139–S142. https://doi.org/10.1016/j.forsciint.2004.09.042

Budowle, B., & Van Daal, A. (2008). Forensically relevant SNP classes. *Biotechniques, 44*(5), 603–610. https://doi.org/10.2144/000112806

Calafell, F., Anglada, R., Bonet, N., González-Ruiz, M., Prats-Muñoz, G., Rasal, R., Lalueza-Fox, C., Bertranpetit, J., Malgosa, A., & Casals, F. (2016). An assessment of a massively parallel sequencing approach for the identification of individuals from mass graves of the Spanish Civil War (1936–1939). *Electrophoresis, 37*(21), 2841–2847. https://doi.org/10.1002/elps.201600180

Casals, F., Anglada, R., Bonet, N., Rasal, R., van der Gaag, K. J., Hoogenboom, J., Solé-Morata, N., Comas, D., & Calafell, F. (2017). Length and repeat-sequence variation in 58 STRs and 94 SNPs in two Spanish populations. *Forensic Science International: Genetics, 30*, 66–70. https://doi.org/10.1016/j.fsigen.2017.06.006

Chaitanya, L., Pajnič, I. Z., Walsh, S., Balažic, J., Zupanc, T., & Kayser, M. (2017). Bringing colour back after 70 years: Predicting eye and hair colour from skeletal remains of World War II victims using the HIrisPlex system. *Forensic Science International: Genetics, 26*, 48–57. https://doi.org/10.1016/j.fsigen.2016.10.004

Cha, S., Lim, J. E., Park, A. Y., Do, J. H., Lee, S. W., Shin, C., Cho, N. H., Kang, J. O., Nam, J. M., Kim, J. S., Woo, K. M., Lee, S. H., Kim, J. Y., & Oh, B. (2018). Identification of five novel genetic loci related to facial morphology by genome-wide association studies. *BMC Genomics, 19*(1). https://doi.org/10.1186/s12864-018-4865-9

Churchill, J. D., Novroski, N. M. M., King, J. L., Seah, L. H., & Budowle, B. (2017). Population and performance analyses of four major populations with Illumina's FGx forensic genomics system. *Forensic Science International: Genetics, 30*, 81–92. https://doi.org/10.1016/j.fsigen.2017.06.004

Claes, P., Liberton, D. K., Daniels, K., Rosana, K. M., Quillen, E. E., Pearson, L. N., McEvoy, B., Bauchet, M., Zaidi, A. A., Yao, W., Tang, H., Barsh, G. S., Absher, D. M., Puts, D. A., Rocha, J., Beleza, S., Pereira, R. W., Baynam, G., Suetens, P., & Shriver, M. D. (2014). Modeling 3D facial shape from DNA. *PLoS Genetics, 10*(3). https://doi.org/10.1371/journal.pgen.1004224

Claes, P., Roosenboom, J., White, J. D., Swigut, T., Sero, D., Li, J., Lee, M. K., Zaidi, A., Mattern, B. C., Liebowitz, C., Pearson, L., González, T., Leslie, E. J., Carlson, J. C., Orlova, E., Suetens, P., Vandermeulen, D., Feingold, E., Marazita, M. L., & Weinberg, S. M. (2018). Genome-wide mapping of global-to-local genetic effects on human facial shape. *Nature Genetics, 50*(3), 414–423. https://doi.org/10.1038/s41588-018-0057-4

Cobain, I. (2016). Killer breakthrough: The day DNA evidence first nailed a murderer. *The Guardian, 7*.

Cole, J., Manyama, M., Kimwaga, E., Mathayo, J., Larson, J., Liberton, D., Lukowiak, K., Ferrara, M., Riccardi, S., Li, M., Prochazkova, M., Williams, T., Jones, K., Klein, O., Santorico, S., Hallgrimsson, B., & Spritz, R. A. (2016). Genome-wide association study of African children identifies association of SCHIP1 and PDE8A with facial size and shape. *PLoS Genetics, 12*(8). https://doi.org/10.1371/journal.pgen.1006174. PMID: 27560698.

Dario, P., Mouriño, H., Oliveira, A. R., Lucas, I., Ribeiro, T., Porto, M. J., Costa Santos, J., Dias, D., & Corte Real, F. (2015). Assessment of IrisPlex-based multiplex for eye and skin color prediction with application to a Portuguese population. *International Journal of Legal Medicine, 129*(6), 1191–1200. https://doi.org/10.1007/s00414-015-1248-5

Davenport, G. C., & Davenport, C. B. (1907). Heredity of eye-color in man. *Science, 26*(670), 589–592. https://doi.org/10.1126/science.26.670.589-b

Devesse, L., Ballard, D., Davenport, L., Riethorst, I., Mason-Buck, G., & Syndercombe Court, D. (2018). Concordance of the ForenSeq™ system and characterisation of sequence-specific autosomal STR alleles across two major population groups. *Forensic Science International: Genetics, 34*, 57–61. https://doi.org/10.1016/j.fsigen.2017.10.012

Donnelly, M. P., Paschou, P., Grigorenko, E., Gurwitz, D., Barta, C., Lu, R. B., Zhukova, O. V., Kim, J. J., Siniscalco, M., New, M., Li, H., Kajuna, S. L. B., Manolopoulos, V. G., Speed, W. C., Pakstis, A. J., Kidd, J. R., & Kidd, K. K. (2012). A global view of the OCA2-HERC2 region and pigmentation. *Human Genetics, 131*(5), 683–696. https://doi.org/10.1007/s00439-011-1110-x

Draus-Barini, J., Walsh, S., Pośpiech, E., Kupiec, T., Głab, H., Branicki, W., & Kayser, M. (2013). Bona fide colour: DNA prediction of human eye and hair colour from ancient and contemporary skeletal remains. *Investigative Genetics, 4*(1). https://doi.org/10.1186/2041-2223-4-3

Duffy, D. L., Box, N. F., Chen, W., Palmer, J. S., Montgomery, G. W., James, M. R., Hayward, N. K., Martin, N. G., & Sturm, R. A. (2004). Interactive effects of MC1R and OCA2 on melanoma risk phenotypes. *Human Molecular Genetics, 13*(4), 447–461. https://doi.org/10.1093/hmg/ddh043

Duffy, D. L., Montgomery, G. W., Chen, W., Zhao, Z. Z., Le, L., James, M. R., Hayward, N. K., Martin, N. G., & Sturm, R. A. (2007). A three-single-nucleotide polymorphism haplotype in intron 1 of OCA2 explains most human eye-color variation. *The American Journal of Human Genetics, 80*(2), 241–252. https://doi.org/10.1086/510885

Durrett, R., & Limic, V. (2001). On the quantity and quality of single nucleotide polymorphisms in the human genome. *Stochastic Processes and their Applications, 93*(1), 1–24. https://doi.org/10.1016/S0304-4149(00)00090-9

Edwards, A., Civitello, A., Hammond, H. A., & Caskey, C. T. (1991). DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. *The American Journal of Human Genetics, 49*(4), 746–756.

Edwards, A., Hammond, H. A., Jin, L., Caskey, C. T., & Chakraborty, R. (1992). Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. *Genomics, 12*(2), 241–253. https://doi.org/10.1016/0888-7543(92)90371-X

Fujimoto, A., Kimura, R., Ohashi, J., Omi, K., Yuliwulandari, R., Batubara, L., Mustofa, M. S., Samakkarn, U., Settheetham-ishida, W., Ishida, T., Morishita, Y., Furusawa, T., Nakazawa, M., Ohtsuka, R., & Tokunaga, K. (2008). A scan for genetic determinants of human hair morphology: EDAR is associated with Asian hair thickness. *Human Molecular Genetics, 17*(6), 835–843. https://doi.org/10.1093/hmg/ddm355

Fujimoto, A., Nishida, N., Kimura, R., Miyagawa, T., Yuliwulandari, R., Batubara, L., Mustofa, M. S., Samakkarn, U., Settheetham-Ishida, W., Ishida, T., Morishita, Y., Tsunoda, T., Tokunaga, K., & Ohashi, J. (2009). FGFR2 is associated with hair thickness in Asian populations. *Journal of Human Genetics, 54*(8), 461–465. https://doi.org/10.1038/jhg.2009.61

Fullwiley, D. (2008). Can DNA 'witness' race?: Forensic uses of an imperfect ancestry testing technology. *GeneWatch, 21*, 12–14.

Fullwood, M. J., Wei, C. L., Liu, E. T., & Ruan, Y. (2009). Next-generation DNA sequencing of paired-end tags (PET) for transcriptome and genome analyses. *Genome Research, 19*(4), 521–532. https://doi.org/10.1101/gr.074906.107

Gill, P. (2001). An assessment of the utility of single nucleotide polymorphisms (SNPs) for forensic purposes. *International Journal of Legal Medicine, 114*(4–5), 204–210. https://doi.org/10.1007/s004149900117

Gužvić, M. (2013). The history of DNA sequencing. *Journal of Medical Biochemistry, 32*(4), 301–312. https://doi.org/10.2478/jomb-2014-0004

Haeusler, M., Haas, C., Lösch, S., Moghaddam, N., Villa, I. M., Walsh, S., Kayser, M., Seiler, R., Ruehli, F., Janosa, M., & Papageorgopoulou, C. (2016). Multidisciplinary identification of the controversial freedom fighter Jörg Jenatsch, assassinated 1639 in Chur, Switzerland. *PLoS One, 11*(12). https://doi.org/10.1371/journal.pone.0168014

Haidar, M., Abbas, F. A., Alsaleh, H., & Haddrill, P. R. (2021). Population genetics and forensic utility of 23 autosomal PowerPlex fusion 6C STR loci in the Kuwaiti population. *Scientific Reports, 11*(1). https://doi.org/10.1038/s41598-021-81425-y

Haidar, M., Alsaleh, H., & Haddrill, P. R. (2020). Population genetics of 30 insertion/deletion polymorphisms in the Kuwaiti population. *International Journal of Legal Medicine, 134*(3), 985–986. https://doi.org/10.1007/s00414-019-02180-4

Hood, L. E., Hunkapiller, M. W., & Smith, L. M. (1987). Automated DNA sequencing and analysis of the human genome. *Genomics, 1*(3), 201–212. https://doi.org/10.1016/0888-7543(87)90046-2

Iozzi, S., Carboni, I., Contini, E., Pescucci, C., Frusconi, S., Nutini, A. L., Torricelli, F., & Ricci, U. (2015). Forensic genetics in NGS era: New frontiers for massively parallel typing. *Forensic Science International: Genetics Supplement Series, 5*, e418–e419. https://doi.org/10.1016/j.fsigss.2015.09.166

Jablonski, N. G., & Chaplin, G. (2010). Human skin pigmentation as an adaptation to UV radiation. *Proceedings of the National Academy of Sciences, 107*(Suppl. 2), 8962–8968. https://doi.org/10.1073/pnas.0914628107

Jeffreys, A. J., Brookfield, J. F. Y., & Semeonoff, R. (1985). Positive identification of an immigration test-case using human DNA fingerprints. *Nature, 317*(6040), 818–819. https://doi.org/10.1038/317818a0

Jeffreys, A. J., Wilson, V., & Thein, S. L. (1985). Individual-specific "fingerprints" of human DNA. *Nature, 316*(6023), 76–79. https://doi.org/10.1038/316076a0

Kanetsky, P. A., Ge, F., Najarian, D., Swoyer, J., Panossian, S., Schuchter, L., Holmes, R., Guerry, D. P., & Rebbeck, T. R. (2004). Assessment of polymorphic variants in the melanocortin-1 receptor gene with cutaneous pigmentation using an evolutionary approach. *Cancer Epidemiology, Biomarkers & Prevention, 13*(5), 808–819.

Kayser, M. (2015). Forensic DNA Phenotyping: Predicting human appearance from crime scene material for investigative purposes. *Forensic Science International: Genetics, 18*, 33–48. https://doi.org/10.1016/j.fsigen.2015.02.003

Kayser, M., Liu, F., Janssens, A. C. J. W., Rivadeneira, F., Lao, O., van Duijn, K., Vermeulen, M., Arp, P., Jhamai, M. M., van IJcken, W. F. J., den Dunnen, J. T., Heath, S., Zelenika, D., Despriet, D. D. G., Klaver, C. C. W., Vingerling, J. R., de Jong, P. T. V. M., Hofman, A., Aulchenko, Y. S., & van Duijn, C. M. (2008). Three genome-wide association studies and a linkage analysis identify HERC2 as a human Iris color gene. *The American Journal of Human Genetics, 82*(2), 411–423. https://doi.org/10.1016/j.ajhg.2007.10.003

Keating, B., Bansal, A. T., Walsh, S., Millman, J., Newman, J., Kidd, K., Budowle, B., Eisenberg, A., Donfack, J., Gasparini, P., Budimlija, Z., Henders, A. K., Chandrupatla, H., Duffy, D. L., Gordon, S. D., Hysi, P., Liu, F., Medland, S. E., Rubin, L., & Kayser, M. (2013). First all-in-one diagnostic tool for DNA intelligence: Genome-wide inference of biogeographic ancestry, appearance, relatedness, and sex with the identitas v1 forensic chip. *International Journal of Legal Medicine, 127*(3), 559–572. https://doi.org/10.1007/s00414-012-0788-1

Kidd, K. K., Pakstis, A. J., Speed, W. C., Lagacé, R., Chang, J., Wootton, S., Haigh, E., & Kidd, J. R. (2014). Current sequencing technology makes microhaplotypes a powerful new type of genetic marker for forensics. *Forensic Science International: Genetics, 12*, 215–224. https://doi.org/10.1016/j.fsigen.2014.06.014

Kidd, K. K., Pakstis, A. J., Speed, W. C., Lagace, R., Wootton, S., & Chang, J. (2018). Selecting microhaplotypes optimized for different purposes. *Electrophoresis, 39*(21), 2815–2823. https://doi.org/10.1002/elps.201800092

Kim, S. Y., Lee, H. C., Chung, U., Ham, S. K., Lee, H. Y., Park, S. J., Roh, Y. J., & Lee, S. H. (2018). Massive parallel sequencing of short tandem repeats in the Korean population. *Electrophoresis, 39*(21), 2702–2707. https://doi.org/10.1002/elps.201800090

King, J. L., Churchill, J. D., Novroski, N. M. M., Zeng, X., Warshauer, D. H., Seah, L. H., & Budowle, B. (2018). Increasing the discrimination power of ancestry- and identity-informative SNP loci within the ForenSeq™ DNA Signature Prep Kit. *Forensic Science International: Genetics, 36*, 60–76. https://doi.org/10.1016/j.fsigen.2018.06.005

King, T. E., Fortes, G. G., Balaresque, P., Thomas, M. G., Balding, D., Delser, P. M., Neumann, R., Parson, W., Knapp, M., Walsh, S., Tonasso, L., Holt, J., Kayser, M., Appleby, J., Forster, P., Ekserdjian, D., Hofreiter, M., & Schürer, K. (2014). Identification of the remains of king richard III. *Nature Communications, 5*. https://doi.org/10.1038/ncomms6631

Köcher, S., Müller, P., Berger, B., Bodner, M., Parson, W., Roewer, L., & Willuweit, S. (2018). Interlaboratory validation study of the ForenSeq™ DNA Signature Prep Kit. *Forensic Science International: Genetics, 36*, 77–85. https://doi.org/10.1016/j.fsigen.2018.05.007

Kostrzewa, G., Konarzewska, M., & Pepiński, W. (2017). Application of massively parallel sequencing (MPS) in paternity testing – case report. *Archives of Forensic Medicine and Criminology, 1*(1), 61–67. https://doi.org/10.5114/amsik.2017.70338

Krier, J. B., Kalia, S. S., & Green, R. C. (2016). Genomic sequencing in clinical practice: Applications, challenges, and opportunities. *Dialogues in Clinical Neuroscience, 18*(3), 299–312. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5067147/pdf/DialoguesClinNeurosci-18-299.pdf.

Lamason, R. L., Mohideen, M. A. P. K., Mest, J. R., Wong, A. C., Norton, H. L., Aros, M. C., Jurynec, M. J., Mao, X., Humphreville, V. R., Humbert, J. E., Sinha, S., Moore, J. L., Jagadeeswaran, P., Zhao, W., Ning, G., Makalowska, I., McKeigue, P. M., O'Donnell, D., Kittles, R., & Cheng, K. C. (2005). Genetics: SLC24A5, a putative cation exchanger, affects pigmentation in zebrafish and humans. *Science, 310*(5755), 1782–1786. https://doi.org/10.1126/science.1116238

LaRue, B. L., Ge, J., King, J. L., & Budowle, B. (2012). A validation study of the Qiagen Investigator DIPplex® kit; an INDEL-based assay for human identification. *International Journal of Legal Medicine, 126*(4), 533–540. https://doi.org/10.1007/s00414-012-0667-9

Lee, M. K., Shaffer, J. R., Leslie, E. J., Orlova, E., Carlson, J. C., Feingold, E., Marazita, M. L., & Weinberg, S. M. (2017). Genome-wide association study of facial morphology reveals novel associations with FREM1 and PARK2. *PLoS One, 12*(4). https://doi.org/10.1371/journal.pone.0176566

Liu, F., Hendriks, A. E. J., Ralf, A., Boot, A. M., Benyi, E., Sävendahl, L., Oostra, B. A., Van Duijn, C., Hofman, A., Rivadeneira, F., Uitterlinden, A. G., Drop, S. L. S., & Kayser, M. (2014). Common DNA variants predict tall stature in Europeans. *Human Genetics, 133*(5), 587–597. https://doi.org/10.1007/s00439-013-1394-0

Liu, F., van Duijn, K., Vingerling, J. R., Hofman, A., Uitterlinden, A. G., Janssens, A. C. J. W., & Kayser, M. (2009). Eye color and the prediction of complex phenotypes from genotypes. *Current Biology, 19*(5), R192–R193. https://doi.org/10.1016/j.cub.2009.01.027

Liu, F., van der Lijn, F., Schurmann, C., Zhu, G., Chakravarty, M. M., Hysi, P. G., Wollstein, A., Lao, O., de Bruijne, M., Ikram, M. A., van der Lugt, A., Rivadeneira, F., Uitterlinden, A. G., Hofman, A., Niessen, W. J., Homuth, G., de Zubicaray, G., McMahon, K. L., Thompson, P. M., & Gibson, G. (2012). A genome-wide association study identifies five loci influencing facial morphology in Europeans. *PLoS Genetics, 8*(9), e1002932. https://doi.org/10.1371/journal.pgen.1002932

Ma, Y., Kuang, J. Z., Nie, T. G., Zhu, W., & Yang, Z. (2016). Next generation sequencing: Improved resolution for paternal/maternal duos analysis. *Forensic Science International: Genetics, 24*, 83–85. https://doi.org/10.1016/j.fsigen.2016.05.015

Marciniak, S., & Perry, G. H. (2017). Harnessing ancient genomes to study the history of human adaptation. *Nature Reviews Genetics, 18*(11), 659–674. https://doi.org/10.1038/nrg.2017.65

Maxam, A. M., & Gilbert, W. (1977). A new method for sequencing DNA. *Proceedings of the National Academy of Sciences of the United States of America, 74*(2), 560–564. https://doi.org/10.1073/pnas.74.2.560

McCarthy, J. J., McLeod, H. L., & Ginsburg, G. S. (2013). Genomic medicine: A decade of successes, challenges, and opportunities. *Science Translational Medicine, 5*(189). https://doi.org/10.1126/scitranslmed.3005785

Medland, S. E., Nyholt, D. R., Painter, J. N., McEvoy, B. P., McRae, A. F., Zhu, G., Gordon, S. D., Ferreira, M. A. R., Wright, M. J., Henders, A. K., Campbell, M. J., Duffy, D. L., Hansell, N. K., Macgregor, S., Slutske, W. S., Heath, A. C., Montgomery, G. W., & Martin, N. G. (2009). Common variants in the trichohyalin gene are associated with straight hair in Europeans. *The American Journal of Human Genetics, 85*(5), 750–755. https://doi.org/10.1016/j.ajhg.2009.10.009

Metzker, M. L. (2010). Sequencing technologies the next generation. *Nature Reviews Genetics, 11*(1), 31–46. https://doi.org/10.1038/nrg2626

Motro, Y., & Moran-Gilad, J. (2017). Next-generation sequencing applications in clinical bacteriology. *Biomolecular Detection and Quantification, 14*, 1–6. https://doi.org/10.1016/j.bdq.2017.10.002

Novroski, N. M. M., King, J. L., Churchill, J. D., Seah, L. H., & Budowle, B. (2016). Characterization of genetic sequence variation of 58 STR loci in four major population groups. *Forensic Science International: Genetics, 25*, 214–226. https://doi.org/10.1016/j.fsigen.2016.09.007

Oldoni, F., Kidd, K. K., & Podini, D. (2019). Microhaplotypes in forensic genetics. *Forensic Science International: Genetics, 38*, 54–69. https://doi.org/10.1016/j.fsigen.2018.09.009

Pakstis, A. J., Speed, W. C., Fang, R., Hyland, F. C. L., Furtado, M. R., Kidd, J. R., & Kidd, K. K. (2010). SNPs for a universal individual identification panel. *Human Genetics, 127*(3), 315–324. https://doi.org/10.1007/s00439-009-0771-1

Palencia-Madrid, L., & de Pancorbo, M. M. (2015). Next generation sequencing as a suitable analysis for authentication of degraded and ancient DNA. *Forensic Science International: Genetics Supplement Series, 5*, e338–e340. https://doi.org/10.1016/j.fsigss.2015.09.134

Paternoster, L., Zhurov, A. I., Toma, A. M., Kemp, J. P., Pourcain, B., St., Timpson, N. J., McMahon, G., McArdle, W., Ring, S. M., Smith, G. D., Richmond, S., & Evans, D. M. (2012). Genome-wide association study of three-dimensional facial morphology identifies a variant in PAX3 associated with nasion position. *The American Journal of Human Genetics, 90*(3), 478–485. https://doi.org/10.1016/j.ajhg.2011.12.021

Phillips, C., Prieto, L., Fondevila, M., Salas, A., Gómez-Tato, A., Álvarez-Dios, J., Alonso, A., Blanco-Verea, A., Brión, M., Montesino, M., Carracedo, A., & Lareu, M. V. (2009). Ancestry analysis in the 11-M Madrid bomb attack investigation. *PLoS One, 4*(8). https://doi.org/10.1371/journal.pone.0006583

Phillips, C., Salas, A., Sánchez, J. J., Fondevila, M., Gómez-Tato, A., Álvarez-Dios, J., Calaza, M., de Cal, M. C., Ballard, D., Lareu, M. V., & Carracedo, Á. (2007). Inferring ancestral origin using a single multiplex assay of ancestry-informative marker SNPs. *Forensic Science International: Genetics, 1*(3–4), 273–280. https://doi.org/10.1016/j.fsigen.2007.06.008

Poinar, H. N., Schwarz, C., Qi, J., Shapiro, B., MacPhee, R. D. E., Buigues, B., Tikhonov, A., Huson, D. M., Tomsho, L. P., Auch, A., Rampp, M., Miller, W., & Schuster, S. C. (2006). Metagenomics to paleogenomics: Large-scale sequencing of mammoth DNA. *Science, 311*(5759), 392–394. https://doi.org/10.1126/science.1123360

Pośpiech, E., Chen, Y., Kukla-Bartoszek, M., Breslin, K., Aliferi, A., Andersen, J. D., Ballard, D., Chaitanya, L., Freire-Aradas, A., van der Gaag, K. J., Girón-Santamaría, L., Gross, T. E., Gysi, M., Huber, G., Mosquera-Miguel, A., Muralidharan, C., Skowron, M., Carracedo, Á., Haas, C., & Kayser, M. (2018). Towards broadening forensic DNA Phenotyping beyond pigmentation: Improving the prediction of head hair shape from DNA. *Forensic Science International: Genetics, 37*, 241–251. https://doi.org/10.1016/j.fsigen.2018.08.017

Pośpiech, E., Wojas-Pelc, A., Walsh, S., Liu, F., Maeda, H., Ishikawa, T., Skowron, M., Kayser, M., & Branicki, W. (2014). The common occurrence of epistasis in the determination of human pigmentation and its impact on DNA-based pigmentation phenotype prediction. *Forensic Science International: Genetics, 11*(1), 64–72. https://doi.org/10.1016/j.fsigen.2014.01.012

Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., & Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics, 38*(8), 904–909. https://doi.org/10.1038/ng1847

Purps, J., Geppert, M., Nagy, M., & Roewer, L. (2011). Evaluation of the IrisPlex eye colour prediction tool in a German population sample. *Forensic Science International: Genetics Supplement Series, 3*(1), e202–e203. https://doi.org/10.1016/j.fsigss.2011.08.101

Ramani, A., Wong, Y., Tan, S. Z., Shue, B. H., & Syn, C. (2017). Ancestry prediction in Singapore population samples using the Illumina ForenSeq kit. *Forensic Science International: Genetics, 31*, 171–179. https://doi.org/10.1016/j.fsigen.2017.08.013

Raymond, P. W., Velie, B. D., & Wade, C. M. (2022). Forensic DNA phenotyping: *Canis familiaris* breed classification and skeletal phenotype prediction using functionally significant skeletal SNPs and indels. *Animal Genetics, 53*(3), 247–263. https://doi.org/10.1111/age.13165

Redler, S., Brockschmidt, F. F., Tazi-Ahnini, R., Drichel, D., Birch, M. P., Dobson, K., Giehl, K. A., Herms, S., Refke, M., Kluck, N., Kruse, R., Lutz, G., Wolff, H., Böhm, M., Becker, T., Nöthen, M. M., Messenger, A. G., & Betz, R. C. (2012). Investigation of the male pattern baldness major genetic susceptibility loci AR/EDA2R and 20p11 in female pattern hair loss. *British Journal of Dermatology, 166*(6), 1314–1318. https://doi.org/10.1111/j.1365-2133.2012.10877.x

Rees, J. L. (2003). Genetics of hair and skin color. *Annual Review of Genetics, 37*, 67–90. https://doi.org/10.1146/annurev.genet.37.110801.143233

Reich, D. E., Schaffner, S. F., Daly, M. J., McVean, G., Mullikin, J. C., Higgins, J. M., Richter, D. J., Lander, E. S., & Altshuler, D. (2002). Human genome sequence variation and the influence of gene history, mutation and recombination. *Nature Genetics, 32*(1), 135–142. https://doi.org/10.1038/ng947

Richards, J. B., Yuan, X., Geller, F., Waterworth, D., Bataille, V., Glass, D., Song, K., Waeber, G., Vollenweider, P., Aben, K. K. H., Kiemeney, L. A., Walters, B., Soranzo, N., Thorsteinsdottir, U., Kong, A., Rafnar, T., Deloukas, P., Sulem, P., Stefansson, H., & Mooser, V. (2008). Male-pattern baldness susceptibility locus at 20p11. *Nature Genetics, 40*(11), 1282–1284. https://doi.org/10.1038/ng.255

Salvoro, C., Faccinetto, C., Zucchelli, L., Porto, M., Marino, A., Occhi, G., de los Campos, G., & Vazza, G. (2019). Performance of four models for eye color prediction in an Italian population sample. *Forensic Science International: Genetics, 40*, 192–200. https://doi.org/10.1016/j.fsigen.2019.03.008

Sanchez, J. J., Phillips, C., Børsting, C., Balogh, K., Bogus, M., Fondevila, M., Harrison, C. D., Musgrave-Brown, E., Salas, A., Syndercombe-Court, D., Schneider, P. M., Carracedo, A., & Morling, N. (2006). A multiplex assay with 52 single nucleotide polymorphisms for human identification. *Electrophoresis, 27*(9), 1713–1724. https://doi.org/10.1002/elps.200500671

Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America, 74*(12), 5463–5467. https://doi.org/10.1073/pnas.74.12.5463

Schneider, P. M., Prainsack, B., & Kayser, M. (2019). The use of forensic DNA phenotyping in predicting appearance and biogeographic ancestry. *Deutsches Arzteblatt International, 116*(51–52), 873–880. https://doi.org/10.3238/arztebl.2019.0873

Shaffer, J. R., Orlova, E., Lee, M. K., Leslie, E. J., Raffensperger, Z. D., Heike, C. L., Cunningham, M. L., Hecht, J. T., Kau, C. H., Nidey, N. L., Moreno, L. M., Wehby, G. L., Murray, J. C., Laurie, C. A., Laurie, C. C., Cole, J., Ferrara, T., Santorico, S., Klein, O., & Weinberg, S. M. (2016). Genome-wide association study reveals multiple loci influencing normal human facial morphology. *PLoS Genetics, 12*(8). https://doi.org/10.1371/journal.pgen.1006149

Sheehan, M. J., & Nachman, M. W. (2014). Morphological and population genomic evidence that human faces have evolved to signal individual identity. *Nature Communications, 5*. https://doi.org/10.1038/ncomms5800

Shendure, J., Balasubramanian, S., Church, G. M., Gilbert, W., Rogers, J., Schloss, J. A., & Waterston, R. H. (2017). DNA sequencing at 40: Past, present and future. *Nature, 550*(7676). https://doi.org/10.1038/nature24286

Silvia, A. L., Shugarts, N., & Smith, J. (2017). A preliminary assessment of the ForenSeq™ FGx system: Next generation sequencing of an STR and SNP multiplex. *International Journal of Legal Medicine, 131*(1), 73–86. https://doi.org/10.1007/s00414-016-1457-6

Smith, L. M., Fung, S., Hunkapiller, M. W., Hunkapiller, T. J., & Hood, L. E. (1985). The synthesis of oligonucleotides containing an aliphatic amino group at the 5′ terminus: Synthesis of fluorescent DNA primers for use in DNA sequence analysis. *Nucleic Acids Research, 13*(7), 2399–2412. https://doi.org/10.1093/nar/13.7.2399

Smith, L. M., Sanders, J. Z., Kaiser, R. J., Hughes, P., Dodd, C., Connell, C. R., Heiner, C., Kent, S. B. H., & Hood, L. E. (1986). Fluorescence detection in automated DNA sequence analysis. *Nature, 321*(6071), 674–679. https://doi.org/10.1038/321674a0

Soejima, M., & Koda, Y. (2007). Population differences of two coding SNPs in pigmentation-related genes SLC24A5 and SLC45A2. *International Journal of Legal Medicine, 121*(1), 36–39. https://doi.org/10.1007/s00414-006-0112-z

Sturm, R. A., Duffy, D. L., Zhao, Z. Z., Leite, F. P. N., Stark, M. S., Hayward, N. K., Martin, N. G., & Montgomery, G. W. (2008). A single SNP in an evolutionary conserved region within intron 86 of the HERC2 gene determines human blue-brown eye color. *The American Journal of Human Genetics, 82*(2), 424–431. https://doi.org/10.1016/j.ajhg.2007.11.005

Sturm, R. A., & Larsson, M. (2009). Genetics of human iris colour and patterns. *Pigment Cell and Melanoma Research, 22*(5), 544–562. https://doi.org/10.1111/j.1755-148X.2009.00606.x

Valentine, T., & Fitzgerald, R. J. (2016). Identifying the culprit: An international perspective on the national academy of sciences report on eyewitness identification evidence. *Applied Cognitive Psychology, 30*(1), 135–138. https://doi.org/10.1002/acp.3164

van Daal, A. (2008). The genetic basis of human pigmentation. *Forensic Science International: Genetics Supplement Series, 1*(1), 541–543. https://doi.org/10.1016/j.fsigss.2007.10.011

Valenzuela, R. K., Henderson, M. S., Walsh, M. H., Garrison, N. A., Kelch, J. T., Cohen-Barak, O., Erickson, D. T., John Meaney, F., Bruce Walsh, J., Cheng, K. C., Ito, S., Wakamatsu, K., Frudakis, T., Thomas, M., & Brilliant, M. H. (2010). Predicting phenotype from genotype: Normal pigmentation. *Journal of Forensic Sciences, 55*(2), 315–322. https://doi.org/10.1111/j.1556-4029.2009.01317.x

Walsh, S., Chaitanya, L., Clarisse, L., Wirken, L., Draus-Barini, J., Kovatsi, L., Maeda, H., Ishikawa, T., Sijen, T., De Knijff, P., Branicki, W., Liu, F., & Kayser, M. (2014). Developmental validation of the HIrisPlex system: DNA-based eye and hair colour prediction for forensic and anthropological usage. *Forensic Science International: Genetics, 9*(1), 150–161. https://doi.org/10.1016/j.fsigen.2013.12.006

Walsh, S., Lindenbergh, A., Zuniga, S. B., Sijen, T., De Knijff, P., Kayser, M., & Ballantyne, K. N. (2011). Developmental validation of the IrisPlex system: Determination of blue and brown iris colour for forensic intelligence. *Forensic Science International: Genetics, 5*(5), 464–471. https://doi.org/10.1016/j.fsigen.2010.09.008

Walsh, S., Liu, F., Wollstein, A., Kovatsi, L., Ralf, A., Kosiniak-Kamysz, A., Branicki, W., & Kayser, M. (2013). The HIrisPlex system for simultaneous prediction of hair and eye colour from DNA. *Forensic Science International: Genetics, 7*(1), 98–115. https://doi.org/10.1016/j.fsigen.2012.07.005

Walsh, S., Wollstein, A., Liu, F., Chakravarthy, U., Rahu, M., Seland, J. H., Soubrane, G., Tomazzoli, L., Topouzis, F., Vingerling, J. R., Vioque, J., Fletcher, A. E., Ballantyne, K. N., & Kayser, M. (2012). DNA-based eye colour prediction across Europe with the IrisPlex system. *Forensic Science International: Genetics, 6*(3), 330–340. https://doi.org/10.1016/j.fsigen.2011.07.009

Wang, J., Li, W., Zhou, N., Liu, J., Zhang, S., Li, X., Li, Z., Yang, Z., Sun, M., & Li, M. (2021). Identification of a novel mutation in the KITLG gene in a Chinese family with familial progressive hyper- and hypopigmentation. *BMC Medical Genomics, 14*(1). https://doi.org/10.1186/s12920-020-00851-5

Weber-Lehmann, J., Schilling, E., Gradl, G., Richter, D. C., Wiehler, J., & Rolf, B. (2014). Finding the needle in the haystack: Differentiating /identical/ twins in paternity testing and forensics by ultra-deep next generation sequencing. *Forensic Science International: Genetics, 9*(1), 42–46. https://doi.org/10.1016/j.fsigen.2013.10.015

Weber, J. L., & Wong, C. (1993). Mutation of human short tandem repeats. *Human Molecular Genetics, 2*(8), 1123–1128. https://doi.org/10.1093/hmg/2.8.1123

Wendt, F. R., Churchill, J. D., Novroski, N. M. M., King, J. L., Ng, J., Oldt, R. F., McCulloh, K. L., Weise, J. A., Smith, D. G., Kanthaswamy, S., & Budowle, B. (2016). Genetic analysis of the yavapai native Americans from west-central Arizona using the illumina MiSeq FGx™ forensic genomics system. *Forensic Science International: Genetics, 24*, 18–23. https://doi.org/10.1016/j.fsigen.2016.05.008

Wray, N. R., & Goddard, M. E. (2010). Multi-locus models of genetic risk of disease. *Genome Medicine, 2*(2). https://doi.org/10.1186/gm131

Xavier, C., & Parson, W. (2017). Evaluation of the illumina ForenSeq™ DNA signature prep kit – MPS forensic application for the MiSeq FGx™ benchtop sequencer. *Forensic Science International: Genetics, 28*, 188–194. https://doi.org/10.1016/j.fsigen.2017.02.018

Zhang, Q., Zhou, Z., Liu, Q., Liu, L., Shao, L., Zhang, M., Ding, X., Gao, Y., & Wang, S. (2018). Evaluation of the performance of illumina's ForenSeq™ system on serially degraded samples. *Electrophoresis, 39*(21), 2674–2684. https://doi.org/10.1002/elps.201800101

CHAPTER 16

Forensic DNA phenotyping in the next-generation sequencing era

Thássia Mayra Telles Carratto[1], Maria Luiza Guimarães de Oliveira[2] and Celso Teixeira Mendes-Junior[1]

[1]Departamento de Química, Laboratório de Pesquisas Forenses e Genômicas, Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, São Paulo, Brazil; [2]Departamento de Genética, Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, São Paulo, Brazil

Introduction

When a biological sample is recovered from a crime scene, obtaining the genetic profile of this individual using autosomal Short Tandem Repeats (STRs) markers on a capillary electrophoresis (CE) platform is the standard approach implemented in investigative forensic laboratories. It was introduced in forensic casework in 1994 and since then it has been considered as the "gold standard" of genetic identification. This technique allowed the progress of forensic DNA analysis and the subsequent creation of national DNA databases. The first database was launched in 1995 by the British Home Office, and in 1997, the Austrian DNA database of the Ministry of the Interior became effective (Parson, 2018). The USA's National DNA Index System (NDIS) was implemented soon after, in 1998.

The main goal of STR profiling is to determine if a person could be the donor of a sample recovered from the crime scene. This is done by comparing the genetic profiles of the query sample and reference samples from victims and suspects. If there is no match, the unidentified profile can be compared with the ones included in DNA databases. Similarly, in missing person or mass disaster cases, the genetic profiles obtained from the victims can be compared to profiles from missing persons' relatives or reference samples collected from victims' items. However, in any of these situations, the failure to identify the unknown profile may slow down or cease the investigative process. If there are no additional suspects or leads, the crime could turn into a cold case.

Therefore, traditional DNA profiling, although extremely valuable, may not be immediately helpful in specific situations (Kayser & Schneider, 2009). Due to occasional limitations in the use of STR profiles, other approaches for the use of DNA for forensic analysis have begun to be investigated.

Next Generation Sequencing (NGS) Technology in DNA Analysis
ISBN 978-0-323-99144-5, https://doi.org/10.1016/B978-0-323-99144-5.00033-0
© 2024 Elsevier Inc.
All rights reserved.

Forensic DNA Phenotyping

In the late 2000s, a new and promising area of study emerged: Forensic DNA Phenotyping (FDP). This methodology aims to predict externally visible characteristics (EVCs) of an individual based solely on their DNA, obtained from the biological material left behind at the crime scene, or from unknown bodies, for example. Thus, this technique would provide additional leads that could help the police to continue investigating the case or narrow down the list of suspects (Kayser, 2015).

The nomenclature "Forensic DNA Phenotyping" first appeared in 2008, in an article published by Koops and Schellekens (2008) that discussed the regulatory issues of the FDP technique in forensic casework, i.e., to what extent it should be allowed. However, the idea of inferring physical traits of an individual for helping police investigations already existed, under the name "DNA-based prediction." Therefore, studies aiming to understand the molecular basis of some EVCs were already being conducted. Such studies were carried out searching for associations between a given physical trait and SNPs of genes that could be involved in the arising of that phenotype, through gene-based or Genome-Wide Association Studies (GWAS). Usually, GWAS are limited to variation sites genotyped through array chips. In the following sections, we will discuss the advantages brought by Next-Generation Sequencing (NGS) to FDP analysis.

An SNP-based tool

A positive aspect of FDP is that appearance traits are inferred mostly through Single Nucleotide Polymorphisms (SNPs) analysis. In forensic casework, SNPs present some advantages over STRs. Since a SNP is defined by the change of only one nucleotide in the DNA sequence, very short PCR amplicons are necessary for its accurate genotyping (e.g., 50 nucleotides or less). In forensic casework, it is not unusual that the sample recovered is in low amounts and/or degraded, rendering it impossible to genotype STRs due to the presence of large alleles composed of many repeats of a given sequence. Therefore, the short amplicon size required for SNP analysis increases the chance of a successful analysis of the recovered DNA. Moreover, the alleles of a SNP are easily designated due to the absence of a repetitive structure, which may produce artifacts such as stutter peaks that sometimes challenge STRs analyses (Kayser & De Knijff, 2011).

The main disadvantage of SNPs for human identification purposes is that SNPs are mostly biallelic, and then they are less polymorphic than multiallelic STRs. Thus, they present limited heterozygosity and are less informative (Wendt & Novroski, 2019). Despite that, this type of marker is abundant and is spread throughout the whole genome. The 1000 Genomes Project has already identified more than 80 million SNPs with minor allele frequency greater than 1%. They also estimated that the differences between a typical and a reference human genome can vary from ~3.5 million to ~4.3 million SNPs (The 1000 Genomes Consortium, 2015). Other studies have concluded that the

analysis of 50–100 autosomal SNPs can provide match probabilities comparable to the 10–15 STRs used for forensic purposes (Pakstis et al., 2010; Sanchez et al., 2006). Nowadays, the multiplex analysis of large amounts of SNPs is not a problem anymore due to the NGS technology, also known as massively parallel sequencing (MPS). This methodology has improved considerably over the past decade. While its cost and effort time have decreased, its sensitivity and ease of use have increased. All these aspects together are facilitating the access of NGS technology to genetics laboratories worldwide. A survey conducted in 105 laboratories (from police, governmental, academic, and private institutions) of 32 European countries revealed that 46% of the participants already owned an MPS platform, and 26.7% have ordered it or were planning to do so within the next 2 years. Since data were collected during 2019, it is expected that, nowadays, 73% of the labs interviewed have an operational MPS platform. MiSeq/MiSeqFGx (Illumina/Verogen) and Ion PGM/Ion S5 (Thermo Fisher Scientific) are the most used platforms in those labs (Gross et al., 2021).

For prediction of complex traits

All EVCs are considered polygenic and complex traits, which means that other factors (internal or external) besides the DNA sequence influence human physical traits. Skin color, for example, is primarily determined by the type and amount of melanin present in its tissue, which is mainly regulated by the genetic background of an individual. As an external factor influencing skin pigmentation, we can cite the sunlight effect: skin color can change if it is exposed to sunlight (and has some degree of tanning ability), becoming darker. As an internal factor, hormonal differences between males and females, or even between individuals of the same sex, may influence the skin pigmentation outcome (Slominski et al., 2004). Thus, this phenotype may vary significantly irrespectively of the genotypic background (Serre et al., 2018).

The same scenario holds in different ways for other EVCs, such as hair color. Some people experience a darkening in this trait as long as they pass from childhood to adulthood (e.g., light blond hair as a child and dark blond/light brown hair as an adult). The molecular mechanism that explains this phenomenon is not fully understood; however, studies have shown that changes in the morphology of a melanosome with age (e.g., its enlargement), which can be caused by modified expression of genes involved in melanogenesis, can be related with hair color changes (Commo et al., 2012; Itou, 2018). Some EVCs are less complicated than others; as the complexity of a trait increases, the difficulty of predicting it accurately increases as well.

Human pigmentation as a starting point

When the development of the FDP area began, the molecular basis of many EVCs was already being studied; however, human pigmentation was the trait that had its genetic

base most well understood at the time. It is considered a less intricate trait when compared to other physical traits. Pigmentation is primarily determined by the presence of melanin in the epidermis, iris, and hair. As many diseases or inflammatory processes are related to changes in pigmentation, the melanin biosynthesis pathway was already under scrutiny before the emergence of FDP. In 2007, the first GWAS on human pigmentation traits was published. Afterward, many other GWAS for eye, hair, and skin color were performed. Nowadays, it is estimated that more than 100 genes are involved in human pigmentation, and *HERC2*, *OCA2*, *TYR*, *MC1R*, *SLC45A2*, *SLC24A5*, and *TYRP1* genes have already been identified as important contributors to this trait (Liu et al., 2013; Pavan & Sturm, 2019).

In 2011, the first tested and validated FDP tool was developed for eye color prediction: the IrisPlex system, based on six highly informative SNPs from six genes. It classifies eye color into three categories: blue, intermediate, and brown. The tool relies on a multinomial linear regression (MLR) algorithm to predict blue and brown eyes with over 90% sensitivity (Walsh, Liu, et al., 2011). Later, in 2013, the HIrisPlex system was launched, including 24 markers (23 SNPs and 1 indel) for eye and hair color simultaneous prediction (Walsh et al., 2013). It consists of the well-known IrisPlex system in addition to another MLR model for hair color prediction based on 22 markers (4 SNPs are used in both models). Hair color is classified into four categories: blond, brown, black, and red. HIrisPlex's sensitivity ranges from 33.3% for black to 66.5% for blond hair. Then, in 2018, an MLR model based on 36 SNPs for skin color prediction was developed and combined with the other two predictive models, conceiving the HIrisPlex-S system (Chaitanya et al., 2018; Walsh et al., 2017). Comprising 41 markers, this tool was proposed for the simultaneous prediction of eye, hair, and skin color. Skin color categories were determined according to the Fitzpatrick scale, which considers the skin tone and the ability to tan. There are five categories: very pale, pale, intermediate, dark, and dark-to-black skin. Its sensitivity for skin color ranges from 7.8% for very pale skin to 81.0% for dark-to-black skin. The three predictive models are freely available online and all their quality metrics are displayed on the tool's website (https://hirisplex.erasmusmc.nl/).

Despite being currently the most well-established tool for eye color prediction, IrisPlex presents a critical limitation since it struggles with intermediate eyes (i.e., non-blue/non-brown colors). This fact is recognized by its developers, and the sensitivity displayed at the HIrisPlex-S official website for this trait is only 0.1%. The model has been tested in populations of different ancestry backgrounds, and the results are consistent: the tool did not produce almost any intermediate eye outcome (Balanovska et al., 2020; Carratto et al., 2021; Dario et al., 2015; Dembinski & Picard, 2014; Kastelic et al., 2013; Meyer et al., 2021; Palmal et al., 2021; Pneuman et al., 2012; Salvoro et al., 2019; Yun et al., 2014). However, when the MLR prediction model is adjusted based on the allele frequencies of the tested population, intermediate eye color sensitivity increases,

ranging from 10% to 65% (Dario et al., 2015; Dembinski & Picard, 2014; Meyer et al., 2021; Palmal et al., 2021; Salvoro et al., 2019).

There are fewer studies evaluating hair and skin prediction models. HIrisPlex-S, when tested on 6500 Latin Americans from the Candela Consortium, presented AUC (Area Under the Receiver Operating Characteristic Curve) values ranging from 65% for brown to 90% for blond hair, and from 71% for intermediate to 78% for light skin. The results for hair are close to the ones obtained for a Brazilian population sample: AUC values ranged from 70% for brown to 88% for blond hair (Carratto et al., 2021). When tested in North Eurasian populations, AUC values between 76% (brown hair) and 92% (black hair) were observed (Balanovska et al., 2020).

Another set of predictive tools is hosted at the Snipper app suite 3 website (http://mathgene.usc.es/snipper/). This portal was designed by the Bioinformatics Group of the Faculty of Mathematics of the University of Santiago de Compostela and hosts different classification tools for human populations. For eye color prediction, the user can choose between a set of 7, 13, or 23 SNPs, and three classification algorithms: naïve-Bayes classifier, MLR, or genetic distance. Eye color can be classified into three categories: blue, green/hazel, and brown (Ruiz et al., 2013). For hair color prediction, a set of 12 SNPs and the same three classification algorithms is available for classifying the phenotype in blond, brown, black, or red color (Söchtig et al., 2015). For skin prediction, the user also can choose between those three classification algorithms to predict its color into white, intermediate, or black, using 10 SNPs (Maroñas et al., 2014). Like the HIrisPlex-S system, the Snipper tools are user-friendly and available online at no cost.

HIrisPlex-S and Snipper are the best known tools, but there are other predictive models in the literature, like the one proposed by Hart et al. for eye and skin color prediction, relying on genotypes of 8 SNPs (Hart et al., 2013), and the models proposed by Allwood and Harbison (2013), which use three and four SNPs to classify eye color in blue, intermediate and brown. More recently, further sets were developed. The EyeColour 11 (EC11) SNP set was proposed by Meyer et al. for eye color prediction in Scandinavians (Meyer et al., 2021). Palmal et al. (2021) proposed the CAN set, including 56, 101, and 120 SNPs for eye, hair, and skin color prediction, respectively. This set was designed based on the GWAS data from over 6500 Latin Americans of the CANDELA consortium.

Of all models cited above, the HIrisPlex-S system is the only one with a validated low throughput assay. Along with IrisPlex evolution, multiplex assays for genotyping the required markers were also designed, based on single base extension (SBE)— SNaPshot — and CE. The first assay for the 6 IrisPlex SNPs showed great sensitivity, producing complete profiles from 31 pg of input DNA. This is extremely important for forensic casework since samples are often found degraded or in low amounts (Walsh, Lindenbergh, et al., 2011). When HIrisPlex was launched a few years later, a new multiplex assay was designed containing its 24 markers. The minimum amount of DNA required for analysis was set at 63 pg, which continues to be an adequate quantity for forensic

casework (Walsh et al., 2014). Due to the limited multiplex capacity of SNapShot (around 30 markers), the 41 markers of the most recent HIrisPlex-S system had to be split into two batches. Therefore, Chaitanya et al. maintained the previous HIrisPlex assay with 24 markers and developed an additional one with the 17 remaining SNPs. The new assay also preserved the sensitivity of 63 pg to produce complete genotypes. Thus, the HIrisPlex-S system comprises two multiplex assays, each one requiring 63 pg of DNA, and three statistical prediction models for eye, hair, and skin color inference (Chaitanya et al., 2018).

HIrisPlex-S multiplex assays were developed based on the SNaPshot and CE technologies aiming its use for police investigations, since basically every investigative laboratory of forensic genetics owns these instruments. However, the fact that the SNaPshot genotyping assay cannot handle a large number of DNA variants is an important disadvantage for the FDP technique. Since this area is still in development, new predictive markers are still being identified and, therefore, the existing tools will probably be updated in the future to include such markers. Using SNaPshot/CE technologies, more multiplex assays would have to be designed to include new markers. It would considerably increase the amount of DNA required, as well as the time, cost, and effort employed in the analysis. In addition, it is noteworthy that FDP analysis would usually be carried out only after the traditional STR analysis; therefore the amount of DNA left for FDP analysis would be decreased. Also, the CE methodology presents limitations for SNP analysis when there is a DNA mixture, which could be very usual in rape cases, for example, due to its semi-quantitative dye-based signals.

Considerations regarding the applicability of NGS to forensic DNA phenotyping

The use of NGS technology overcomes the previously mentioned limitations. It has an increased multiplex capacity, allowing the analysis of thousands of markers in a single assay. Thus, envisaging the transition from CE to MPS, the HIrisPlex-S developer group proposed a protocol to generate and analyze its 41 markers from raw MPS data (Breslin et al., 2019). They also performed the forensic developmental validation of the MPS assay on two platforms commonly used in forensics: Ion Torrent and MiSeq. Samples simulating casework were obtained from blood, and other sources, such as saliva, hair bulbs, semen, and low quantity touch DNA. Using commercial control samples in serial concentrations (5, 10, 25, 50, 100, 250, 500 pg, and 1 ng), the Ion Torrent MPS assay was able to produce complete genetic profiles from 100 pg of DNA, while the MiSeq MPS assay required 250 pg of DNA. This input difference may be due to the unequal number of samples used in each run. While MiSeq sequenced 96 samples with one cartridge, Ion Torrent sequenced the 96 samples split into two chips with 48 samples each. Therefore, if a reduced number of samples were analyzed in MiSeq, the sensitivity achieved for this platform could have been higher. In fact, except for touch DNA, both platforms yielded

complete profiles for all sources simulating casework from 100 pg of DNA. Although this minimum amount of required DNA is higher than the amount required for the two HIrisPlex-S SNaPshot assays (63 pg), in the MPS platform all the markers can be analyzed in a single run, consuming DNA only once. Currently, the 24 HIrisPlex markers for eye and color simultaneous prediction are included in the ForenSeq DNA Signature Prep Kit (Verogen) for preparing samples' library to be sequenced on MiSeq (Illumina) and MiSeq FGx (Verogen) platforms. HIrisPlex markers also compose the Ion AmpliSeq DNA Phenotyping Panel, which is compatible with the Applied Bioystems Precision ID sequencing workflow and can be sequenced on Thermo Fisher platforms.

In the same study, a protocol to interpret a mixture of two samples from NGS data was proposed. Forty mixtures were simulated and an average of 28 HIrisPlex-S markers could be separated into two profiles. An evaluation of sensitivity for artificially degraded samples prepared for sequencing on the MiSeq platform was performed. 500 pg of DNA was exposed to Ultraviolet (UV) radiation through different time windows. Complete HIrisPlex-S genetic profiles were obtained even after 10 min of UV light exposure, establishing MPS platforms as a robust methodology for forensic purposes.

In spite of the importance of HIrisPlex-S in the FDP area, its accuracy can still be improved, even for Europeans. In such populations, pigmentation traits may show some degree of correlation, such as the common cooccurrence of blue eyes and light skin, or dark skin and brown eyes. A study with individuals from Europe and the United States estimated and added the correlation between pigmentation phenotypes to the HIrisPlex-S prediction model. They observed that: (a) the accuracy of the system increased, and (b) the 41 DNA variants could explain only a fraction of this correlation. Therefore, they conclude that more independent predictive markers need to be identified, and their addition to the current established predictive tools would increase its accuracy (Chen et al., 2021). Pigmentation traits (as well as other EVCs) are determined by variation sites of major effects, in addition to variation sites of minor effects. Thus, work is needed to continue unraveling the genetic basis of human pigmentation. In 2021, 50 and 111 new variants influencing eye and hair color, respectively, were identified in two large GWAS conducted on thousands of Europeans (Hysi et al., 2018; Simcoe et al., 2021).

Moreover, to assure the global performance of predictive tools, the accomplishment of GWAS in non-European populations is extremely important, with particular attention to admixed populations. GWAS conducted on African populations have revealed novel aspects of their genetic architecture (Crawford et al., 2017; Martin et al., 2017). Regarding admixed populations, Latin Americans present a complex admixture pattern involving Native Americans, Europeans, and sub-Saharan Africans: Amerindian ancestry is stronger in Mexico, Guatemala, Peru, and Ecuador, while Cuba, Chile, Colombia, Puerto Rico, Venezuela, Argentina, Brazil, and Uruguay have a substantial contribution from Europeans, and countries in the Caribbean area (such as Haiti and Jamaica) display higher levels of African ancestry (Norris et al., 2018; Rodrigues-Soares et al., 2020;

Ruiz-Linares et al., 2014; Salzano & Sans, 2014). Due to the admixed background, non-European alleles (or combinations of them) derived from different biogeographical ancestries may be influencing pigmentation in such populations as well. For example, recent studies have identified associations between the *MFSD12* region and skin color in East Asians and Native Americans, supporting the theory that light skin has evolved independently in West and East Eurasia (Adhikari et al., 2019; Lona-Durazo et al., 2019).

Massively Parallel Sequencing assays designed to explore the genetic diversity of pigmentation genes, notably in non-European populations, may play a role of utmost importance in identifying less common variation sites with minor effects that are hardly identified by array-based GWAS. Moreover, the need to increase the number of independent markers coupled with the need of introducing non-European variation sites into forthcoming FDP tools in order to achieve more accurate eye, hair, and skin color predictions may restrict their use to Next-Generation Sequencing assays. And this conclusion may be easily extended to all polygenic EVCs.

Other EVCs as FDP targets

We have discussed only pigmentation traits so far because they already have established predictive tools that were tested in many different populations, highlighting their strengths and weaknesses that must be improved. However, there are many other promising EVCs currently being studied.

It has been demonstrated that many loci associated with human pigmentation are also associated with the presence of freckles (Eriksson et al., 2010; Sulem et al., 2007, 2008), and some predictive models have already been developed. The first one was published in 2018 and used SNPs from four genes (*MC1R*, *IRF4*, *ASIP,* and *BNC2*) to classify individuals as freckled or nonfreckled, achieving an accuracy of 74.13% (Hernando et al., 2018). Another study developed two models based on 17 and 19 SNPs: a binary (i.e., freckled/nonfreckled) and a multinomial (non/medium/heavily freckled) one. The models also considered gender and interaction between SNPs as predictive variables. The binary model achieved a sensitivity of 84.0%, while the sensitivity of the multinomial was 54.5%, 74.3%, and 17.2% for non, medium, and heavily freckled individuals, respectively (Kukla-Bartoszek et al., 2019).

Body height is a characteristic that has been studied since the beginning of the FDP era, because it is commonly measured in many cohort studies for different purposes and, therefore, this information was promptly available. It has been estimated that heritability for human height is around 80% (Liu et al., 2019). However, height is a complex and continuous trait, and it is estimated that hundreds or thousands of SNPs are required for an accurate prediction. Since the exact stature is very hard to predict, some models have been developed converting the stature into a discrete variable using binary classifications (i.e., tall − nontall). Three GWAS were conducted by the International Genetics

of Anthropometric Traits (GIANT) (Allen et al., 2010; Wood et al., 2014; Yengo et al., 2018). The first one identified 180 SNPs associated with height, while the second study highlighted 697 new SNPs and the third, 512 new SNPs. Liu et al. investigated the predictive power of the first two sets in Europeans and achieved AUC values of 0.75 and 0.79, respectively (Liu et al., 2019). Lello et al. (2018) conducted a study for age prediction in Europeans using approximately 20,000 SNPs and obtained a correlation of 0.65 between the predicted and the actual height of individuals.

The eyebrow color is highly correlated with scalp hair color; however, this correlation is not perfect and, therefore, eyebrow and scalp hair color may differ. This suggests that, in spite of some variants influencing both traits, there are variants unique for eyebrow color. A GWAS conducted in 2019 has identified a new locus associated with eyebrow color (*C10orf11*) and five other loci that are known for their associations with scalp hair color. From these six loci, the authors developed a model using 25 SNPs for eyebrow color prediction and obtained AUC values ranging from 62.0% for brown to 70.1% for blond eyebrows (Peng, Zhu, et al., 2019).

Genes involved in the molecular basis of hair shape have been highlighted during the past decade. The first two genes reported as associated with this trait were *EDAR* and *TCHH*. While *EDAR* has a major effect on explaining straight hair in East Asians, the *TCHH* gene plays a major role in Europeans. *GATA3*, *PRSS53*, *FRAS1*, and *WNT10A* genes have shown associations with hair shape as well (Adhikari et al., 2016; Pospiech et al., 2015). Liu et al. developed a prediction model for straight versus non-straight hair using 14 SNPs (Liu et al., 2018), which achieved AUCs of 0.66 and 0.64 in two different cohorts (individuals with North-Western European ancestry from Australia and North-Western Europeans from the Netherlands). After that, Pośpiech et al. developed two models, a binary one, using 32 SNPs to predict straight versus non-straight hair, and a multinomial one, using 33 SNPs to predict straight versus wavy versus curly hair (Pospiech et al., 2015). The binary model presented an AUC of 0.68, while the multinomial model presented AUCs of 0.68, 0.60, and 0.62 for straight, wavy, and curly hair, respectively.

Still discussing hair aspects, it has been shown that the most associated SNPs for male hair loss are located in the *AR/EDA2R* region, on chromosome X. In addition, more than 250 loci related to this condition have been identified so far. Some predictive models have been developed, with sensitivities around 73%–84% (Hagenaars et al., 2017; Liu et al., 2016; Pirastu et al., 2017).

Predicting the facial features of an individual can be considered the ultimate goal of forensic DNA phenotyping. This task is a big challenge since the human face has many different characteristics that must be taken into account. Furthermore, there are many different ways in which the face can be subdivided when looking for associations. In the past decade, some independent GWAS for facial shape, most of them focusing on Europeans, have highlighted different associated SNPs. Ten genes were identified in

more than one study: *CACNA2D3*, *DCHS2*, *EPHB3*, *HOXD*, *PAX1*, *PAX3*, *PKDCC*, *SOX9*, *SUPT3H*, and *TBX15*. A recent GWAS study looking for associations between markers and facial distances in Europeans established 13 facial points and measured the distance between them. They identified 24 genetic loci (17 of them were novel) and then validated these findings in additional multiethnic populations. Ten loci were successfully replicated, six of them being novel (Xiong et al., 2019).

Taken together, the performance parameters observed for all these EVCs highlight the fact that the discovery of SNPs from array-based GWAS studies is of limited extent. Given the polygenic nature of these EVCs, massively parallel sequencing association-based studies focusing on pertinent gene panels and involving populations from multiple ethnic backgrounds is mandatory for real progress in this field. Such studies would not only allow the identification of population-specific variation sites of major effect and minor effect variation sites of worldwide relevance, but also a proper assessment of population-specific linkage disequilibrium within the pertinent genes in order to assist with the selection of independent markers for FDP purposes.

Although biogeographic ancestry (BGA) is not a morphological trait, it is frequently considered under the FDP scope. So far, many ancestry informative markers (AIMs) have been identified, and many sets for BGA inference have been developed. The *SNPforID* 34-plex forensic ancestry test is one of the most established sets. It includes 34 SNPs and is focused on the differentiation of sub-Saharan Africans, Europeans, and East Asians (Fondevila et al., 2013). The Snipper app suite 3 website hosts an online tool that allows the use of the 34-plex set. To fill the gap of Native American ancestry, the PIMA (Population Informative Multiplex for the Americas) set was published in 2020. It could be considered as a complement of the previous 34-plex, and comprises 26 AIM-SNPs, focusing on differentiating American ancestry from other continental populations (Carvalho Gontijo et al., 2020). Currently, there are other established BGA inference sets, such as Seldin's 128 AIMs and the Kidd lab set of 55 AIMs (Kidd et al., 2014; Kosoy et al., 2009).

DNA methylation and age prediction

Even though age cannot be directly predicted from the DNA sequence, since it barely changes over time, epigenetic marks are more compliant with changes throughout a lifetime. In this realm, DNA methylation markers have been suggested as the most informative tool for age prediction (Freire-Aradas et al., 2020). DNA methylation is essential for silencing retroviral elements, regulating tissue-specific gene expression, genomic imprinting, and X chromosome inactivation (Moore et al., 2013). It involves the attachment of a methyl ($-CH_3$) group to the 5′ carbon (C5) of a cytosine linked to a guanine through a phosphate group (CpG). CpG sites are usually methylated and relatively rare across the genome (Deaton & Bird, 2011), except for specific locations called CpG islands (CGIs). CGIs correspond to regions predominantly unmethylated with high CG density

(>55% CG), mainly located at the gene promoter regions (Antequera & Bird, 1993; Bird et al., 1985).

The presence of 5-methylcytosine (5 mC) in CpG islands represses gene expression by regulating chromatin structure and preventing transcription factors from binding to a promoter region. Since these regions are associated with the control of gene expression, CpG islands could exhibit a tissue-specific pattern of DNA methylation (Moore et al., 2013). However, CpG islands associated with transcription start sites rarely show tissue-specific methylation patterns (Meissner et al., 2008), while regions located as far as 2 kb from CpG islands, called CpG islands shores, present evolutionary conserved patterns of tissue-specific methylation (Irizarry et al., 2009).

Methylation patterns are heritable and transmitted with high fidelity during DNA replication (Jones & Liang, 2009). DNA methyltransferase 1 (DNMT1) maintains the DNA methylation pattern in a cell lineage (Hermann et al., 2004). Passive or active mechanisms can remove these DNA methylation patterns. The passive DNA demethylation occurs in the absence of the DNA methylation maintenance machinery. Active DNA demethylation depends on the activity of proteins from the ten-eleven translocation (TET) family: TET1, TET2, and TET3 (Ito et al., 2010). DNMT3A and DNMT3B enzymes, complexed with histone deacetylases (HDACs) and Histone H3 lysine 9 methylation (H3K9me), are responsible for de novo methylation, which occurs in sites with no previous indication of methylation (Epsztejn-Litman et al., 2008).

5 mC corresponds to approximately 1.5% of genomic DNA (Lister et al., 2009). It is highest in the embryo and decreases gradually over time by a process that includes de novo methylation of CpG islands bound to polycomb repressors and demethylation of other genomic sites. Despite tissue-specific rates and speed, these methylation changes occur in all tissues over time (Dor & Cedar, 2018; Maegawa et al., 2010). Thus, by measuring the methylation status derived from a DNA sample of a given tissue, it is possible to predict the individual's age (Horvath, 2013).

Methylation patterns have been widely used to predict age, since they are altered throughout life and are essential for the aging process. Generally speaking, age-associated epigenetic markers comprise locus-specific hypermethylation and global hypomethylation (Xiao et al., 2016). Moreover, the 5 mC levels are tissue-specific, a factor that justifies the importance of assessing the most suitable predictors for each tissue (Horvath, 2013; Parson, 2018). Although many statistical models have been proposed for predicting age in different tissues, most age prediction studies employ blood or buccal samples (Bocklandt et al., 2011; Eipel et al., 2016; Freire-Aradas et al., 2016; Hong et al., 2017; Park & Kim, 2016; Vidaki et al., 2017; Zbieć-Piekarska et al., 2015).

Bocklandt et al. (2011) developed the first epigenetic age predictor model based on 88 CpG sites from saliva samples and achieved an average accuracy of 5.2 years. Many studies have suggested age prediction models based on CpG sites located in the *ELOVL2* gene (Freire-Aradas et al., 2016; Hamano et al., 2016; Hannum et al., 2013; Park & Kim,

2016; Peng, Feng, et al., 2019; Zbieć-Piekarska et al., 2015; Zubakov et al., 2016). Hannum et al. (2013) provided a high accuracy chronological age estimator based on 71 CpG sites, showing a 96% correlation with age and a median absolute deviation of 3.9 years. Horvath (2013) studied 8000 samples covering 51 tissues and identified 353 age-associated CpG sites, with a 96% age correlation and 3.6 years average error. Horvath et al. (2018) developed epigenetic clocks to predict age in specific cell types such as skin, endothelial cells, and buccal cells and made available an online DNAm age calculator (https://dnamage.clockfoundation.org/calculator).

Another freely available predictive tool is also noteworthy. It is hosted at the previously addressed Snipper app suite 3 website (http://mathgene.usc.es/snipper/age_models.php) and uses seven markers to infer age (Freire-Aradas et al., 2016), with a predictor error of approximately 3 years. This is the most suitable tool for use in forensic casework, since they involve a limited number of methylation markers that can be easily evaluated by Sanger sequencing after bisulfite conversion (as explained below). However, the identification of new methylation markers associated with age is still necessary.

There are three main groups of techniques providing the identification of regions that are differentially methylated: bisulfite conversion—based methods, restriction enzyme—based approaches, and affinity enrichment—based assays (Pajares et al., 2021). The restriction enzyme—based methods take advantage of the differential digestion properties of methylation-sensitive restriction enzymes (MSREs) and the affinity enrichment—based methods employ either methyl-CpG-binding domain (MBD) proteins or antibodies specific for 5 mC (as in MeDIP) to enrich methylated DNA regions. Notwithstanding, bisulfite conversion methods are the standard approach to analyzing methylation patterns. The technique takes advantage of differential deamination kinetics of unmethylated and methylated cytosines. Before PCR amplification, the unmethylated cytosines will be deaminated to uracil, while methylated cytosine remains unaffected (Parson, 2018). Then, these methylation patterns will be analyzed by downstream techniques, such as sequencing-based approaches (Li & Tollefsbol, 2021) or methylation-specific PCR (Herman et al., 1996). The Illumina Infinium HumanMethylation450 (450 K) Beadchip array-based platform is one of the most used methods for characterization of DNA methylation after bisulfite conversion (Solomon et al., 2018). The Illumina Beadchip array provides a user-friendly and cost-effective alternative for methylation pattern analysis; however, there are associated biases, such as the inclusion of probes that interrogate only previously identified CpG sites (Yong et al., 2016).

Therefore, NGS methodology empowers the study of epigenetic variation, offering high resolution data and a more detailed view of its modifications (Meaburn & Schulz, 2012). These technologies allow the interrogation of genomes without prior knowledge of sequence (Barros-Silva et al., 2018). Therefore, the second and third sequencing-

generation methods are helpful for exploratory bisulfite-converted DNA analysis. One of the greatest advantages of NGS approaches is that precise ratios of methylation may be easily established by observing the presence of hundreds or thousands of reads by locus in different cell types. Moreover, these high-throughput deep sequencing methods are fast, reliable, produce terabytes of data, are relatively accessible for genome-wide analysis, and provide resolution at single base-pairs (Laird, 2010; Soto et al., 2016; van den Oord et al., 2013). This single-base resolution may be of paramount importance for the identification of new age-related methylation sites.

The forensic field has made few attempts at MPS implementation in DNA methylation analysis with age prediction purposes. Vidaki et al. employed a combined method based on machine learning and NGS to generate a model for age prediction from blood samples. Based on identifying the methylation status of 16 CpG sites, they proposed a model achieving an age correlation of 0.86 and a mean absolute error (MAE) of 7.45 years (Vidaki et al., 2017). Naue et al. presented an age-prediction tool based on MPS and a random forest machine learning algorithm. Using whole blood samples, they proposed a model based on four genes (*ELOVL2*, *F5*, *KLF14*, and *TRIM59*) with a MAE of 3.2 years and a root-square error (RMSE) of 4.19 years (Naue et al., 2017).

A model for predicting chronological age combining MPS and machine learning methods was proposed from 12 CpG sites previously selected by Vidaki et al. The model showed an RMSE of 4.9 years with an MAE of 4.1 years (Aliferi et al., 2018). Using the MiSeq FGx instrument (Verogen on Illumina technology), Heidegger et al. introduced the VISAGE primary prototype tool for age estimation aiming at 32 CpG sites from five genes (*ELOVL2*, *MIR29B2C*, *FHL2*, *TRIM59*, and *KLF14*). They observed a mean-standard deviation of 1.4% across ratios for age prediction estimates on blood samples (Heidegger et al., 2020). Following this, Heidegger et al. evaluated the methylation pattern in 13 CpG sites from semen samples. The proposed tool was independently validated by five laboratories from the **Vis**ible **A**ttributes Through **Ge**nomics (VISAGE) project consortium, with the assay performance showing an MAE of 5.1 years (Heidegger et al., 2022). Considering blood, buccal swab, and bone samples, Woźniak proposed three age prediction models comprising the VISAGE Consortiumś enhanced tool for epigenetic age estimation. They achieved a model based on six CpGs from six genes with an MAE of 3.2 years for blood, a model for buccal cells using five CpGs from five genes yielding an MAE of 3.7 years, and a model employing six CpGs from four genes that showed an MAE of 3.4 years for bone samples (Woźniak et al., 2021).

Although NGS approaches are expected to be widely used in the future, this is not yet the method of choice in the case of small-scale projects requiring low-resolution sequencing information regarding methylation, especially due to the complex data analysis and higher cost of data generation (Barros-Silva et al., 2018; Rauluseviciute et al., 2019; Wreczycka et al., 2017).

The VISAGE initiative and the future of FDP

In 2017, the **Vis**ible **A**ttributes Through **Ge**nomics (VISAGE) Consortium was established. It was composed of 13 participants from academic, police, and justice institutions of eight European countries, and from different areas of expertise (forensic geneticists, forensic DNA practitioners, statistical geneticists, and social scientists). The project aimed to further explore FDP technology, conducting more studies about markers associated with appearance, age, and ancestry, and using them to develop new predictive tools. They also focused on the statistical interpretation of data, and how this methodology could be used by investigative authorities. The legal, ethical, and social aspects concerning the implementation of the FDP technique in forensic casework were also discussed.

The project formally came to its end in 2021 and, so far, more than 20 scientific papers comprising all the topics mentioned above have been published. The Consortium has developed two types of predictive tools: a basic and an enhanced one. They were developed aiming at sequencing assays performed on NGS platforms, given that the simultaneous analysis of hundreds of markers will be needed for their successful application. Separate tools were developed for (a) appearance traits (AT) and BGA, and (b) age prediction, since samples must be prepared in different ways for the latter purpose.

The basic tool for AT and BGA prediction uses the 41 HIrisPlex-S markers for eye, hair, and color inference and 115 ancestry informative markers (AIMs) for BGA inference, comprising 153 markers (three of them overlap for BGA and AT prediction). This tool has been validated in two different platforms: a) using the PowerSeq Chemistry on the MiSeq FGx System and b) using a single multiplex reaction using the AmpliSeq design pipeline for the Ion S5 MPS platform. In general, complete genotypic profiles were obtained from as little as 100−125 pg of DNA (Palencia-Madrid et al., 2020; Xavier et al., 2020). The 153 markers are covered by the Ion AmpliSeq VISAGE-Basic Tool Research Panel, suitable for sequencing on ThermoFisher platforms. The enhanced tool for AT and BGA prediction includes 524 markers for predicting eye, hair, skin, and eyebrow color, hair morphology, male baldness, and presence of freckles. BGA inference is performed using autosomal, X, and Y-linked SNPs, as well as microhaplotypes. This tool was validated on Ion S5 platform and tested in five different laboratories, emerging as a robust, sensitive, and reproducible assay for forensic puporses (Xavier et al., 2022).

The VISAGE Consortium has also developed tools for age estimation, as already mentioned in the previous topic. The basic tool is only suitable for blood samples and has been validated using bisulfite conversion and posterior sequencing on the MiSeq FGx platform (Heidegger et al., 2020). The enhanced tool for age estimation performs prediction from three different tissues (blood, buccal cells, and bones), using specific markers and statistical models for each tissue. It was validated on the MiSeq FGx platform (Woźniak et al., 2021). A model for age estimation from semen has also been developed and validated (Heidegger et al., 2022; Pisarek et al., 2021).

The progress made in the FDP area since it started is huge. However, according to everything we have discussed so far, there is still a lot of work to be done. Even though genes that play a major role in the understanding of the genetic basis of EVCs have been highlighted, there is still a need to identify markers that play minor roles to achieve an ideal predictive power. Given that EVCs are complex polygenic traits, it is important to capture the variability of all the genes involved. As we commented before, GWAS are usually conducted based on array methodologies. A possible approach is to calculate polygenic scores for appearance traits using thousands/millions of markers of an array panel. However, even in such cases, many important genetic variants for a specific trait could be missing in the array chip. NGS technology has enabled the possibility of deeply sequencing a whole genome, or even specific genes of interest. Custom panels for targeting genes of interest can be designed and for that, there are many currently available kits in the market. HaloPlex (Agilent), SureSelect (Agilent), KAPA (Roche), Lotus (IDT), and AmpliSeq (Illumina) are some of the well-known kits for customized library preparation available. Moreover, SureSelect offers the possibility of creating a customized panel optimized for sequencing on the Ion Proton platform. Lastly, there is also the possibility of amplifying the DNA by in-house multiplex PCR methods and using Nextera (Illumina) for library preparation. Thus, new markers that may not be present in array chips can be analyzed without the need for imputation. Haplotype analysis is also simplified since NGS can provide phased sequence reads. In addition, as commented before, since thousands of markers can be analyzed in a single run, predictive SNPs can be easily incorporated into library preparation kits as soon as they are identified. Lastly, different types of markers can be analyzed in the same run. The ForenSeq DNA Signature Prep Kit (Verogen) panel, for example, includes 59 STRs and 172 SNPs, which allows the STR profiling and AT/BGA prediction to be performed simultaneously if necessary, saving the amount of DNA required.

It should be mentioned that long-read massively parallel sequencing platforms, such as PacBio and Oxford Nanopore, also known as third-generation sequencers, have been available since 2010. However, their major novelty, i.e., the production of reads that usually exceeds 10,000 bp, may result in difficulties with handling degraded and/or low copy number samples. Therefore, second-generation sequencers remain as the best option for FDP purposes aiming at forensic casework.

FDP for forensic anthropology purposes

The robustness of NGS technology in dealing with degraded samples makes it a powerful tool for analyzing samples from ancient and contemporary skeletal remains, which includes the possibility of predicting EVCs from these remains. Kukla-Bartoszek et al. sequenced 63 bone samples at postmortem intervals up to almost 80 years using an Ion AmpliSeq HIrisPlex-S panel on the Ion Proton platform. Complete profiles were

obtained from 49 pg of DNA. The 41 markers were successfully genotyped in 35 samples (55.6%), while 23 samples presented partial profiles. Although it was not possible to predict skin color for any of the 23 samples with partial profiles, eye and hair color were predicted for 20 and 12 of them, respectively (Kukla-Bartoszek et al., 2020).

Another study analyzed two skeletal remains from La Belle, a French ship that sank in the USA in 1986. Although human remains were recovered in 1996, they were only analyzed in 2015. Samples were prepared with ForenSeq DNA Signature Prep Kit (Verogen) and sequenced on MiSeq FGx Desktop Sequencer (Verogen). Partial profiles were generated for both samples: 216 (93.5%) and 154 (66.7%) markers were successfully called for the two samples. Eye and hair color predictions were made using HIrisPlex, and ancestry was also inferred. Since the ship was European, it was surprising that the first sample had predictions suggesting Native American ancestry, dark hair and brown eyes. The authors hypothesized that these remains could belong to a local Karankawa Indian who was scavenging the shipwreck, became trapped, and perished. The second sample, as expected, had a European origin. The pigmentation trait was not predicted due to insufficient data (Ambers et al., 2020).

In the anthropology field, appearance information of ancient samples may help understand archeological and evolutionary aspects involved in the arising of phenotypes. Using whole-genome sequencing data, HIrisPlex-S and Snipper were employed to predict pigmentation phenotypes of 20 contemporary and 22 ancient Native American samples. Predictions were performed for 7 of 22 ancient samples and for all contemporary ones. Outcomes for both groups of Native Americans were consistent. Ancient individuals were predicted to have intermediate/brown eyes, black hair, and intermediate/darker skin pigmentation (Carratto et al., 2020).

Final remarks

The interest in learning or applying predictive markers for forensic casework has been growing over the years. The survey conducted in 105 European forensic laboratories has shown that 26.7% of them have experience with ancestry SNPs analysis; 21.9% have already done appearance SNPs analysis, and only 14.3% have performed methylation analysis. However, 40% of the participants are interested in learning more about methylation; 37.1% desire to do appearance analysis, and 32.4% plan to perform ancestry inference. Between 15% and 21% of the participants said that they have not used predictive markers for analysis because it is not legally allowed in their respective country of residence (Gross et al., 2021). It is important to note that, along with the reliability and improvement of FDP tools, ethical and legal aspects grounding the approval of this technique for forensic casework should be carefully and extensively discussed. Although this is not the scope of this book, these questions may be reviewed in (Nogel et al., 2019; Samuel & Prainsack, 2019; Toom et al., 2016). Most countries do not have

explicit legislation for FDP use but, at the same time, they do not have an implicit prohibition either, since the development of this technique is still recent. In Europe, the first countries to explicitly regulate FDP use were The Netherlands and Slovakia. In Austria, Germany, and Sweden, the technique is allowed, while in countries such as France, Poland, Spain, the United Kingdom, Australia, South Africa, and Brazil it is not allowed and neither implicitly forbidden. In the United States, legislation varies from state to state (Samuel and Prainsack, 2018, 2020).

As mentioned at the beginning of this chapter, NGS technology is starting to become accessible in many forensic laboratories, at least in Europe. We foresee it as a global trend. The advances in NGS technology are allowing the enhancement of the FDP area, not only technically, but also pushing more countries to start discussing its legal regulations for forensic casework.

References

Adhikari, K., Fontanil, T., Cal, S., Mendoza-Revilla, J., Fuentes-Guajardo, M., Chacón-Duque, J. C., Al-Saadi, F., Johansson, J. A., Quinto-Sanchez, M., Acuña-Alonzo, V., Jaramillo, C., Arias, W., Lozano, R. B., Pérez, G. M, Gómez-Valdés, J., Villamil-Ramírez, H., Hunemeier, T., Ramallo, V., Cerqueira, C. C. S. ..., & Ruiz-Linares, A. (2016). A genome-wide association scan in admixed Latin Americans identifies loci influencing facial and scalp hair features. *Nature Communications, 7*. https://doi.org/10.1038/ncomms10815

Adhikari, K., Mendoza-Revilla, J., Sohail, A., Fuentes-Guajardo, M., Lampert, J., Chacón-Duque, J. C., Hurtado, M., Villegas, V., Granja, V., Acuña-Alonzo, V., Jaramillo, C., Arias, W., Lozano, R. B., Everardo, P., Gómez-Valdés, J., Villamil-Ramírez, H., Cerqueira, C. C. S., Hunemeier, T., Ramallo, V. ..., & Ruiz-Linares, A. (2019). A GWAS in Latin Americans highlights the convergent evolution of lighter skin pigmentation in Eurasia. *Nature Communications, 10*(1). https://doi.org/10.1038/s41467-018-08147-0

Aliferi, A., Ballard, D., Gallidabino, M. D., Thurtle, H., Barron, L., & Syndercombe-Court, D. (2018). DNA methylation-based age prediction using massively parallel sequencing data and multiple machine learning models. *Forensic Science International: Genetics, 37*, 215–226. https://doi.org/10.1016/j.fsigen.2018.09.003

Allen, H. L., Estrada, K., Lettre, G., Berndt, S. I., Weedon, M. N., Rivadeneira, F., Willer, C. J., Jackson, A. U., Vedantam, S., Raychaudhuri, S., Ferreira, T., Wood, A. R., Weyant, R. J., Segrè, A. V., Speliotes, E. K., Wheeler, E., Soranzo, N., Park, J. H., Yang, J. ..., & Hirschhorn, J. N. (2010). Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature, 467*(7317), 832–838. https://doi.org/10.1038/nature09410

Allwood, J. S., & Harbison, S. A. (2013). SNP model development for the prediction of eye colour in New Zealand. *Forensic Science International: Genetics,* 7(4), 444–452. https://doi.org/10.1016/j.fsigen.2013.03.005

Ambers, A., Bus, M. M., King, J. L., Jones, B., Durst, J., Bruseth, J. E, Gill-King, H., & Budowle, B. (2020). Forensic genetic investigation of human skeletal remains recovered from the La Belle shipwreck. *Forensic Science International, 306*, 1–10. https://doi.org/10.1016/j.forsciint.2019.110050

Antequera, F., & Bird, A. (1993). Number of CpG islands and genes in human and mouse. *Proceedings of the National Academy of Sciences of the United States of America, 90*(24), 11995–11999. https://doi.org/10.1073/pnas.90.24.11995

Balanovska, E., Lukianova, E., Kagazezheva, J., Maurer, A., Leybova, N., Agdzhoyan, A., Gorin, I., Petrushenko, V., Zhabagin, M., Pylev, V., Kostryukova, E., & Balanovsky, O. (2020). Optimizing the genetic prediction of the eye and hair color for North Eurasian populations. *BMC Genomics, 21*(Suppl. 7), 1–13. https://doi.org/10.1186/s12864-020-06923-1

Barros-Silva, D., Marques, C. J., Henrique, R., & Jerónimo, C. (2018). Profiling DNA methylation based on next-generation sequencing approaches: New insights and clinical applications. *Genes, 9*(9). https://doi.org/10.3390/genes9090429

Bird, A., Taggart, M., Frommer, M., Miller, O. J., & Macleod, D. (1985). A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA. *Cell, 40*(1), 91–99. https://doi.org/10.1016/0092-8674(85)90312-5

Bocklandt, S., Lin, W., Sehl, M. E., Sanchez, F. J., Sinsheimer, J. S., Horvath, S., & Vilain, E. (2011). Epigenetic predictor of age. *PLoS One, 6*(6), e14821. https://doi.org/10.1371/journal.pone.0014821

Breslin, K., Wills, B., Ralf, A., Ventayol Garcia, M., Kukla-Bartoszek, M., Pospiech, E., Freire-Aradas, A., Xavier, C., Ingold, S., de La Puente, M., van der Gaag, K. J., Herrick, N., Haas, C., Parson, W., Phillips, C., Sijen, T., Branicki, W., Walsh, S., & Kayser, M. (2019). HIrisPlex-S system for eye, hair, and skin color prediction from DNA: Massively parallel sequencing solutions for two common forensically used platforms. *Forensic Science International: Genetics, 43*(May), 102152. https://doi.org/10.1016/j.fsigen.2019.102152

Carratto, T. M. T., Marcorin, L., Debortoli, G., Hünemeier, T., Norton, H., Parra, E. J., Castelli, E. C., & Mendes-Junior, C. T. (2020). Insights on hair, skin and eye color of ancient and contemporary Native Americans. *Forensic Science International: Genetics, 48*(May), 102335. https://doi.org/10.1016/j.fsigen.2020.102335

Carratto, T. M. T., Marcorin, L., Valle-Silva, G., Oliveira, M. L. G, Donadi, E. A., Simões, A. L., Castelli, E. C., & Mendes-Junior, C. T. (2021). Prediction of eye and hair pigmentation phenotypes using the HIrisPlex system in a Brazilian admixed population sample. *International Journal of Legal Medicine, 135*(4), 1329–1339. https://doi.org/10.1007/s00414-021-02554-7

Carvalho Gontijo, C., Porras-Hurtado, L. G., Freire-Aradas, A., Fondevila, M., Santos, C., Salas, A., Henao, J., Isaza, C., Beltrán, L., Nogueira Silbiger, V., Castillo, A., Ibarra, A., Moreno Chavez, F., Söchtig, J., Ruiz, Y., Barreto, G., Rondon, F., Zabala, W., Borjas, L. ..., & Phillips, C. (2020). Pima: A population informative multiplex for the Americas. *Forensic Science International: Genetics, 44*(January). https://doi.org/10.1016/j.fsigen.2019.102200

Chaitanya, L., Breslin, K., Zuñiga, S., Wirken, L., Pośpiech, E., Kukla-Bartoszek, M., Sijen, T., de Knijff, P., Liu, F., Branicki, W., Kayser, M., & Walsh, S. (2018). The HIrisPlex-S system for eye, hair and skin colour prediction from DNA: Introduction and forensic developmental validation. *Forensic Science International: Genetics, 35*(April), 123–135. https://doi.org/10.1016/j.fsigen.2018.04.004

Chen, Y., Branicki, W., Walsh, S., Nothnagel, M., Kayser, M., & Liu, F. (2021). The impact of correlations between pigmentation phenotypes and underlying genotypes on genetic prediction of pigmentation traits. *Forensic Science International: Genetics, 50*. https://doi.org/10.1016/j.fsigen.2020.102395

Commo, S., Wakamatsu, K., Lozano, I., Panhard, S., Loussouarn, G., Bernard, B. A., & Ito, S. (2012). Age-dependent changes in eumelanin composition in hairs of various ethnic origins. *International Journal of Cosmetic Science, 34*(1), 102–107. https://doi.org/10.1111/j.1468-2494.2011.00691.x

Crawford, N. G., Kelly, D. E., Hansen, M. E. B., Beltrame, M. H., Fan, S., Bowman, S. L., Jewett, E., Ranciaro, A., Lo, Y., Pfeifer, S. P., Jensen, J. D., Campbell, M. C., Beggs, W., Hormozdiari, F., Mpoloka, S. W., & George, G. (2017). Loci associated with skin pigmentation identified in African populations. *Science, 358*(6365), 1–39. https://doi.org/10.1126/science.aan8433.Loci

Dario, P., Mouriño, H., Oliveira, A. R., Lucas, I., Ribeiro, T., Porto, M. J., Costa Santos, J., Dias, D., & Corte Real, F. (2015). Assessment of IrisPlex-based multiplex for eye and skin color prediction with application to a Portuguese population. *International Journal of Legal Medicine, 129*(6), 1191–1200. https://doi.org/10.1007/s00414-015-1248-5

Deaton, A. M., & Bird, A. (2011). CpG islands and the regulation of transcription. *Genes and Development, 25*(10), 1010–1022. https://doi.org/10.1101/gad.2037511

Dembinski, G. M., & Picard, C. J. (2014). Evaluation of the IrisPlex DNA-based eye color prediction assay in a United States population. *Forensic Science International: Genetics, 9*(1), 111–117. https://doi.org/10.1016/j.fsigen.2013.12.003

Dor, Y., & Cedar, H. (2018). Principles of DNA methylation and their implications for biology and medicine. *Lancet, 392*(10149), 777–786. https://doi.org/10.1016/S0140-6736(18)31268-6

Eipel, M., Mayer, F., Arent, T., Ferreira, M. R. P., Birkhofer, C., Costa, I. G., Timme, S. R., & Wagner, W. (2016). Epigenetic age predictions based on buccal swabs are more precise in combination with cell type-specific DNA methylation signatures. *Aging (Albany NY), 8*(5), 1034−1048. https://doi.org/10.18632/aging.100972

Epsztejn-Litman, S., Feldman, N., Abu-Remaileh, M., Shufaro, Y., Gerson, A., Ueda, J., Deplus, R., Fuks, F., Shinkai, Y., Cedar, H., & Bergman, Y. (2008). De novo DNA methylation promoted by G9a prevents reprogramming of embryonically silenced genes. *Nature Structural and Molecular Biology, 15*(11), 1176−1183. https://doi.org/10.1038/nsmb.1476

Eriksson, N., Macpherson, J. M., Tung, J. Y., Hon, L. S., Naughton, B., Saxonov, S., Avey, L., Wojcicki, A., Pe'er, I., & Mountain, J. (2010). Web-based, participant-driven studies yield novel genetic associations for common traits. *PLoS Genetics, 6*(6), 1−20. https://doi.org/10.1371/journal.pgen.1000993

Fondevila, M., Phillips, C., Santos, C., Freire Aradas, A., Vallone, P. M., Butler, J. M., Lareu, M. V., & Carracedo, A. (2013). Revision of the SNPforID 34-plex forensic ancestry test: Assay enhancements, standard reference sample genotypes and extended population studies. *Forensic Science International: Genetics, 7*(1), 63−74. https://doi.org/10.1016/j.fsigen.2012.06.007

Freire-Aradas, A., Phillips, C., Mosquera-Miguel, A., Girón-Santamaría, L., Gómez-Tato, A., Casares De Cal, M., Álvarez-Dios, J., Ansede-Bermejo, J., Torres-Español, M., Schneider, P. M., Pośpiech, E., Branicki, W., Carracedo, A., & Lareu, M. V. (2016). Development of a methylation marker set for forensic age estimation using analysis of public methylation data and the Agena Bioscience EpiTYPER system. *Forensic Science International: Genetics, 24*, 65−74. https://doi.org/10.1016/j.fsigen.2016.06.005

Freire-Aradas, A., Pośpiech, E., Aliferi, A., Girón-Santamaría, L., Mosquera-Miguel, A., Pisarek, A., Ambroa-Conde, A., Phillips, C., Casares de Cal, M. A., Gómez-Tato, A., Spólnicka, M., Woźniak, A., Álvarez-Dios, J., Ballard, D., Syndercombe-Court, D., Branicki, W., Carracedo, A., & Lareu, M. V. (2020). A comparison of forensic age prediction models using data from four DNA methylation technologies. *Frontiers in Genetics, 11*(August), 1−12. https://doi.org/10.3389/fgene.2020.00932

Gross, T. E., Fleckhaus, J., & Schneider, P. M. (2021). Progress in the implementation of massively parallel sequencing for forensic genetics: Results of a European-wide survey among professional users. *International Journal of Legal Medicine, 135*(4), 1425−1432. https://doi.org/10.1007/s00414-021-02569-0

Hagenaars, S. P., Hill, W. D., Harris, S. E., Ritchie, S. J., Davies, G., Liewald, D. C., Gale, C. R., Porteous, D. J., Deary, I. J., & Marioni, R. E. (2017). Genetic prediction of male pattern baldness. *PLoS Genetics, 13*(2), 1−16. https://doi.org/10.1371/journal.pgen.1006594

Hamano, Y., Manabe, S., Morimoto, C., Fujimoto, S., Ozeki, M., & Tamaki, K. (2016). Forensic age prediction for dead or living samples by use of methylation-sensitive high resolution melting. *Legal Medicine, 21*, 5−10. https://doi.org/10.1016/j.legalmed.2016.05.001

Hannum, G., Guinney, J., Zhao, L., Zhang, L., Hughes, G., Sadda, S., Klotzle, B., Bibikova, M., Fan, J. B., Gao, Y., Deconde, R., Chen, M., Rajapakse, I., Friend, S., Ideker, T., & Zhang, K. (2013). Genome-wide methylation profiles reveal quantitative views of human aging rates. *Molecular Cell, 49*(2), 359−367. https://doi.org/10.1016/j.molcel.2012.10.016

Hart, K. L., Kimura, S. L., Mushailov, V., Budimlija, Z. M., Prinz, M., & Wurmbach, E. (2013). Improved eye- and skin-color prediction based on 8 SNPs. *Croatian Medical Journal, 54*(3), 248−256. https://doi.org/10.3325/cmj.2013.54.248

Heidegger, A., Xavier, C., Niederstätter, H., de la Puente, M., Pośpiech, E., Pisarek, A., Kayser, M., Branicki, W., & Parson, W. (2020). Development and optimization of the VISAGE basic prototype tool for forensic age estimation. *Forensic Science International: Genetics, 48*(April), 102322. https://doi.org/10.1016/j.fsigen.2020.102322

Heidegger, A., Pisarek, A., de la Puente, M., Niederstätter, H., Pośpiech, E., Woźniak, A., Schury, N., Unterländer, M., Sidstedt, M., Junker, K., Ventayol Garcia, M., Laurent, F. X., Ulus, A., Vannier, J., Bastisch, I., Hedman, J., Sijen, T., Branicki, W., Xavier, C., & Parson, W. (2022). Development and inter-laboratory validation of the VISAGE enhanced tool for age estimation from semen using quantitative DNA methylation analysis. *Forensic Science International: Genetics, 56*, 102596. https://doi.org/10.1016/j.fsigen.2021.102596

Herman, J. G., Graff, J. R., Myohanen, S., Nelkin, B. D., & Baylin, S. B. (1996). Methylation-specific PCR: A novel PCR assay for methylation status of CpG islands. *Proceedings of the National Academy of Sciences of the United States of America, 93*(18), 9821–9826. https://doi.org/10.1073/pnas.93.18.9821

Hermann, A., Gowher, H., & Jeltsch, A. (2004). Biochemistry and biology of mammalian DNA methyltransferases. *Cellular and Molecular Life Sciences, 61*(19–20), 2571–2587. https://doi.org/10.1007/s00018-004-4201-1

Hernando, B., Ibañez, M. V., Deserio-Cuesta, J. A., Soria-Navarro, R., Vilar-Sastre, I., & Martinez-Cadenas, C. (2018). Genetic determinants of freckle occurrence in the Spanish population: Towards ephelides prediction from human DNA samples. *Forensic Science International: Genetics, 33*(June 2017), 38–47. https://doi.org/10.1016/j.fsigen.2017.11.013

Hong, S. R., Jung, S. E., Lee, E. H., Shin, K. J., Yang, W. I., & Lee, H. Y. (2017). DNA methylation-based age prediction from saliva: High age predictability by combination of 7 CpG markers. *Forensic Science International: Genetics, 29*, 118–125. https://doi.org/10.1016/j.fsigen.2017.04.006

Horvath, S. (2013). DNA methylation age of human tissues and cell types. *Genome Biology, 14*(10). https://doi.org/10.1186/gb-2013-14-10-r115

Horvath, S., Oshima, J., Martin, G. M., Lu, A. T., Quach, A., Cohen, H., Felton, S., Matsuyama, M., Lowe, D., Kabacik, S., Wilson, J. G., Reiner, A. P., Maierhofer, A., Flunkert, J., Aviv, A., Hou, L., Baccarelli, A. A., Li, Y., Stewart, J. D. ..., & Raj, K. (2018). Epigenetic clock for skin and blood cells applied to Hutchinson Gilford Progeria Syndrome and ex vivo studies. *Aging (Albany NY), 10*(7), 1758–1775. https://doi.org/10.18632/aging.101508

Hysi, P. G., Valdes, A. M., Liu, F., Furlotte, N. A., Evans, D. M., Bataille, V., Visconti, A., Hemani, G., McMahon, G., Ring, S. M., Smith, G. D., Duffy, D. L., Zhu, G., Gordon, S. D., Medland, S. E., Lin, B. D., Willemsen, G., Jan Hottenga, J., Vuckovic, D. ..., & Spector, T. D. (2018). Genome-wide association meta-analysis of individuals of European ancestry identifies new loci explaining a substantial fraction of hair color variation and heritability. *Nature Genetics, 50*(5), 652–656. https://doi.org/10.1038/s41588-018-0100-5

Irizarry, R. A., Ladd-Acosta, C., Wen, B., Wu, Z., Montano, C., Onyango, P., Cui, H., Gabo, K., Rongione, M., Webster, M., Ji, H., Potash, J. B., Sabunciyan, S., & Feinberg, A. P. (2009). The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nature Genetics, 41*(2), 178–186. https://doi.org/10.1038/ng.298

Ito, S., Dalessio, A. C., Taranova, O. V., Hong, K., Sowers, L. C., & Zhang, Y. (2010). Role of tet proteins in 5 mC to 5 hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature, 466*(7310), 1129–1133. https://doi.org/10.1038/nature09303

Itou, T. (2018). Morphological changes in hair melanosomes by aging. *Pigment Cell Melanoma Res, 31*(5), 630–635. https://doi.org/10.1111/pcmr.12697

Jones, P. A., & Liang, G. (2009). Rethinking how DNA methylation patterns are maintained. *Nature Reviews Genetics, 10*(11), 805–811. https://doi.org/10.1038/nrg2651

Kastelic, V., Pośpiech, E., Draus-Barini, J., Branicki, W., & Drobnič, K. (2013). Prediction of eye color in the slovenian population using the irisPlex SNPs. *Croatian Medical Journal, 54*(4), 381–386. https://doi.org/10.3325/cmj.2013.54.381

Kayser, M. (2015). Forensic DNA Phenotyping: Predicting human appearance from crime scene material for investigative purposes. *Forensic Science International: Genetics, 18*, 33–48. https://doi.org/10.1016/j.fsigen.2015.02.003

Kayser, M., & De Knijff, P. (2011). Improving human forensics through advances in genetics, genomics and molecular biology. *Nature Reviews Genetics, 12*(3), 179–192. https://doi.org/10.1038/nrg2952

Kayser, M., & Schneider, P. M. (2009). DNA-based prediction of human externally visible characteristics in forensics: Motivations, scientific challenges, and ethical considerations. *Forensic Science International: Genetics, 3*(3), 154–161. https://doi.org/10.1016/j.fsigen.2009.01.012

Kidd, K. K., Speed, W. C., Pakstis, A. J., Furtado, M. R., Fang, R., Madbouly, A., Maiers, M., Middha, M., Friedlaender, F. R., & Kidd, J. R. (2014). Progress toward an efficient panel of SNPs for ancestry inference. *Forensic Science International: Genetics, 10*(1), 23–32. https://doi.org/10.1016/j.fsigen.2014.01.002

Koops, B.-J., & Schellekens, M. (2008). Forensic DNA phenotyping: Regulatory issues. *Columbia Science and Technology Law Review, 9*, 158–202. https://doi.org/10.7916/stlr.v9i0.3818

Kosoy, R., Nassir, R., Tian, C., White, P. A., Butler, L. M., Kittles, R., Alarcon-riquelme, M. E., Gregersen, P. K., Belmont, J. W., La Vega, F. M. D., & Seldin, M. F. (2009). Ancestry informative marker sets for determining continental origin and admixture proportions in common populations in America. *Human Mutation, 30*(1), 69–78. https://doi.org/10.1002/humu.20822

Kukla-Bartoszek, M., Pośpiech, E., Woźniak, A., Boroń, M., Karłowska-Pik, J., Teisseyre, P., Zubańska, M., Bronikowska, A., Grzybowski, T., Płoski, R., Spólnicka, M., & Branicki, W. (2019). DNA-based predictive models for the presence of freckles. *Forensic Science International: Genetics, 42*(January), 252–259. https://doi.org/10.1016/j.fsigen.2019.07.012

Kukla-Bartoszek, M., Szargut, M., Pośpiech, E., Diepenbroek, M., Zielińska, G., Jarosz, A., Piniewska-Róg, D., Arciszewska, J., Cytacka, S., Spólnicka, M., Branicki, W., & Ossowski, A. (2020). The challenge of predicting human pigmentation traits in degraded bone samples with the MPS-based HIrisPlex-S system. *Forensic Science International: Genetics, 47*(April), 102301. https://doi.org/10.1016/j.fsigen.2020.102301

Laird, P. W. (2010). Principles and challenges of genome-wide DNA methylation analysis. *Nature Reviews Genetics, 11*(3), 191–203. https://doi.org/10.1038/nrg2732

Lello, L., Avery, S. G., Tellier, L., Vazquez, A. I., de los Campos, G., & Hsu, S. D. H. (2018). Accurate genomic prediction of human height. *Genetics, 210*(2), 477–497. https://doi.org/10.1534/genetics.118.301267

Li, S., & Tollefsbol, T. O. (2021). DNA methylation methods: Global DNA methylation and methylomic analyses. *Methods, 187*(June 2020), 28–43. https://doi.org/10.1016/j.ymeth.2020.10.002

Lister, R., Pelizzola, M., Dowen, R. H., Hawkins, R. D., Hon, G., Tonti-Filippini, J., Nery, J. R., Lee, L., Ye, Z., Ngo, Q. M., Edsall, L., Antosiewicz-Bourget, J., Stewart, R., Ruotti, V., Millar, A. H., Thomson, J. A., Ren, B., & Ecker, J. R. (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature, 462*(7271), 315–322. https://doi.org/10.1038/nature08514

Liu, F., Chen, Y., Zhu, G., Hysi, P. G., Wu, S., Adhikari, K., Breslin, K., Pośpiech, E., Hamer, M. A., Peng, F., Muralidharan, C., Acuna-Alonzo, V., Canizales-Quinteros, C., Bedoya, G., Gallo, C., Poletti, G., Rothhammer, F., Bortolini, M. C., Gonzalez-Jose, R. ..., & Kayser, M. (2018). Meta-analysis of genome-wide association studies identifies 8 novel loci involved in shape variation of human head hair. *Human Molecular Genetics, 27*(3), 559–575. https://doi.org/10.1093/hmg/ddx416

Liu, F., Hamer, M. A., Heilmann, S., Herold, C., Moebus, S., Hofman, A., Uitterlinden, A. G., Nöthen, M. M., van Duijn, C. M., Nijsten, T. E. C., & Kayser, M. (2016). Prediction of male-pattern baldness from genotypes. *European Journal of Human Genetics, 24*(6), 895–902. https://doi.org/10.1038/ejhg.2015.220

Liu, F., Wen, B., & Kayser, M. (2013). Colorful DNA polymorphisms in humans. *Seminars in Cell and Developmental Biology, 24*(6–7), 562–575. https://doi.org/10.1016/j.semcdb.2013.03.013

Liu, F., Zhong, K., Jing, X., Uitterlinden, A. G., Hendriks, A. E. J., Drop, S. L. S., & Kayser, M. (2019). Update on the predictability of tall stature from DNA markers in Europeans. *Forensic Science International: Genetics, 42*(May), 8–13. https://doi.org/10.1016/j.fsigen.2019.05.006

Lona-Durazo, F., Hernandez-Pacheco, N., Fan, S., Zhang, T., Choi, J., Kovacs, M. A., Loftus, S. K., Le, P., Edwards, M., Fortes-Lima, C. A., Eng, C., Huntsman, S., Hu, D., Gómez-Cabezas, E. J., Marín-Padrón, L. C., Grauholm, J., Mors, O., Burchard, E. G., Norton, H. L. ..., & Parra, E. J. (2019). Meta-analysis of GWA studies provides new insights on the genetic architecture of skin pigmentation in recently admixed populations. *BMC Genetics, 20*(1), 1–16. https://doi.org/10.1186/s12863-019-0765-5

Maegawa, S., Hinkal, G., Kim, H. S., Shen, L., Zhang, L., Zhang, J., Zhang, N., Liang, S., Donehower, L. A., & Issa, J. P. J. (2010). Widespread and tissue specific age-related DNA methylation changes in mice. *Genome Research, 20*(3), 332–340. https://doi.org/10.1101/gr.096826.109

Maroñas, O., Phillips, C., Söchtig, J., Gomez-Tato, A., Cruz, R., Alvarez-Dios, J., Casares de Cal, M. A., Ruiz, Y., Fondevila, M., Carracedo, A., & Lareu, M. V. (2014). Development of a forensic skin colour

predictive test. *Forensic Science International: Genetics, 13*, 34–44. https://doi.org/10.1016/j.fsigen.2014.06.017

Martin, A. R., Lin, M., Granka, J. M., Myrick, J. W., Liu, X., Sockell, A., Atkinson, E. G., Werely, C. J., Möller, M., Sandhu, M. S., Kingsley, D. M., Hoal, E. G., Liu, X., Daly, M. J., Feldman, M. W., Gignoux, C. R., Bustamante, C. D., & Henn, B. M. (2017). An unexpectedly complex architecture for skin pigmentation in Africans. *Cell, 171*(6), 1340–1353.e14. https://doi.org/10.1016/j.cell.2017.11.015

Meaburn, E., & Schulz, R. (2012). Next generation sequencing in epigenetics: Insights and challenges. *Seminars in Cell and Developmental Biology, 23*(2), 192–199. https://doi.org/10.1016/j.semcdb.2011.10.010

Meissner, A., Mikkelsen, T. S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein, B. E., Nusbaum, C., Jaffe, D. B., Gnirke, A., Jaenisch, R., & Lander, E. S. (2008). Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature, 454*(7205), 766–770. https://doi.org/10.1038/nature07107

Meyer, O. S., Salvo, N. M., Kjærbye, A., Kjersem, M., Andersen, M. M., Sørensen, E., Ullum, H., Janssen, K., Morling, N., Børsting, C., Olsen, G. H., & Andersen, J. D. (2021). Prediction of eye colour in scandinavians using the eyecolour 11 (Ec11) snp set. *Genes, 12*(6). https://doi.org/10.3390/genes12060821

Moore, L. D., Le, T., & Fan, G. (2013). DNA methylation and its basic function. *Neuropsychopharmacology, 38*(1), 23–38. https://doi.org/10.1038/npp.2012.112

Naue, J., Hoefsloot, H. C. J., Mook, O. R. F., Rijlaarsdam-Hoekstra, L., van der Zwalm, M. C. H., Henneman, P., Kloosterman, A. D., & Verschure, P. J. (2017). Chronological age prediction based on DNA methylation: massive parallel sequencing and random forest regression. *Forensic Science International: Genetics, 31*, 19–28. https://doi.org/10.1016/j.fsigen.2017.07.015

Nogel, M., Pádár, Z., Czebe, A., & Kovács, G. (2019). Developing legal regulation of forensic DNA-phenotyping in Hungary. *Forensic Science International: Genetics Suppl Ser, 7*(1), 609–611. https://doi.org/10.1016/j.fsigss.2019.10.109

Norris, E. T., Wang, L., Conley, A. B., Rishishwar, L., Mariño-Ramírez, L., Valderrama-Aguirre, A., & Jordan, I. K. (2018). Genetic ancestry, admixture and health determinants in Latin America. *BMC Genomics, 19*(Suppl. 8). https://doi.org/10.1186/s12864-018-5195-7

Pajares, M. J., Palanca-Ballester, C., Urtasun, R., Alemany-Cosme, E., Lahoz, A., & Sandoval, J. (2021). Methods for analysis of specific DNA methylation status. *Methods, 187*(April 2020), 3–12. https://doi.org/10.1016/j.ymeth.2020.06.021

Pakstis, A. J., Speed, W. C., Fang, R., Hyland, F. C. L., Furtado, M. R., Kidd, J. R., & Kidd, K. K. (2010). SNPs for a universal individual identification panel. *Human Genetics, 127*(3), 315–324. https://doi.org/10.1007/s00439-009-0771-1

Palencia-Madrid, L., Xavier, C., de la Puente, M., Hohoff, C., Phillips, C., Kayser, M., & Parson, W. (2020). Evaluation of the visage basic tool for appearance and ancestry prediction using powerseq chemistry on the miseq fgx system. *Genes, 11*(6), 1. https://doi.org/10.3390/genes11060708

Palmal, S., Adhikari, K., Mendoza-Revilla, J., Fuentes-Guajardo, M., Cerqueira, C. C. S., Bonfante, B., Chacón-Duque, J. C., Sohail, A., Hurtado, M., Villegas, V., Granja, V., Jaramillo, C., Arias, W., Lozano, R. B., Everardo-Martínez, P., Gómez-Valdés, J., Villamil-Ramírez, H., Hünemeier, T., Ramallo, V. ..., & Ruiz-Linares, A. (2021). Prediction of eye, hair and skin colour in Latin Americans. *Forensic Science International: Genetics, 53*(November 2020). https://doi.org/10.1016/j.fsigen.2021.102517

Park, S. T., & Kim, J. (2016). Trends in next-generation sequencing and a new era for whole genome sequencing. *International Journal of Neurology, 20*, 76–83. https://doi.org/10.5213/inj.1632742.371

Parson, W. (2018). Age estimation with DNA: From forensic DNA fingerprinting to forensic (Epi)genomics: A mini-review. *Gerontology, 64*(4), 326–332. https://doi.org/10.1159/000486239

Pavan, W. J., & Sturm, R. A. (2019). The genetics of human skin and hair pigmentation. *Annual Review of Genomics and Human Genetics, 20*, 41–72. https://doi.org/10.1146/annurev-genom-083118-015230

Peng, F., Feng, L., Chen, J., Wang, L., Li, P., Ji, A., Zeng, C., Liu, F., & Li, C. (2019). Validation of methylation-based forensic age estimation in time-series bloodstains on FTA cards and gauze at room

temperature conditions. *Forensic Science International: Genetics, 40*(December 2018), 168–174. https://doi.org/10.1016/j.fsigen.2019.03.006

Peng, F., Zhu, G., Hysi, P. G., Eller, R. J., Chen, Y., Li, Y., Hamer, M. A., Zeng, C., Hopkins, R. L., Jacobus, C. L., Wallace, P. L., Uitterlinden, A. G., Ikram, M. A., Nijsten, T., Duffy, D. L., Medland, S. E., Spector, T. D., Walsh, S., Martin, N. G., Liu, F., & Kayser, M. (2019). Genome-wide association studies identify multiple genetic loci influencing eyebrow color variation in Europeans. *Journal of Investigative Dermatology, 139*(7), 1601–1605. https://doi.org/10.1016/j.jid.2018.12.029

Pirastu, N., Joshi, P. K., De Vries, P. S., Cornelis, M. C., McKeigue, P. M., Keum, N., Franceschini, N., Colombo, M., Giovannucci, E. L., Spiliopoulou, A., Franke, L., North, K. E., Kraft, P., Morrison, A. C., Esko, T., & Wilson, J. F. (2017). GWAS for male-pattern baldness identifies 71 susceptibility loci explaining 38% of the risk. *Nature Communications, 8*(1), 1–9. https://doi.org/10.1038/s41467-017-01490-8

Pisarek, A., Pośpiech, E., Heidegger, A., Xavier, C., Papież, A., Piniewska-Róg, D., Kalamara, V., Potabattula, R., Bochenek, M., Sikora-Polaczek, M., Macur, A., Woźniak, A., Janeczko, J., Phillips, C., Haaf, T., Polaoska, J., Parson, W., Kayser, M., & Branicki, W. (2021). Epigenetic age prediction in semen - marker selection and model development. *Aging (Albany NY), 13*(15), 19145–19164. https://doi.org/10.18632/aging.203399

Pneuman, A., Budimlija, Z. M., Caragine, T., Prinz, M., & Wurmbach, E. (2012). Verification of eye and skin color predictors in various populations. *Legal Medicine, 14*(2), 78–83. https://doi.org/10.1016/j.legalmed.2011.12.005

Pospiech, E., Karłowska-Pik, J., Marcińska, M., Abidi, S., Andersen, J. D., Van Den Berge, M., Carracedo, A., Eduardoff, M., Freire-Aradas, A., Morling, N., Sijen, T., Skowron, M., Söchtig, J., Syndercombe-Court, D., Weiler, N., Schneider, P. M., Ballard, D., Børsting, C., Parson, W., Phillips, C., & Branicki, W. (2015). Evaluation of the predictive capacity of DNA variants associated with straight hair in Europeans. *Forensic Science International: Genetics, 19*, 280–288. https://doi.org/10.1016/j.fsigen.2015.09.004

Rauluseviciute, I., Drabløs, F., & Rye, M. B. (2019). DNA methylation data by sequencing: Experimental approaches and recommendations for tools and pipelines for data analysis. *Clinical Epigenetics, 11*(1), 1–13. https://doi.org/10.1186/s13148-019-0795-x

Ruiz, Y., Phillips, C., Gomez-Tato, A., Alvarez-Dios, J., Casares De Cal, M. A., Cruz, R., Maroñas, O., Söchtig, J., Fondevila, M., Rodriguez-Cid, M. J., Carracedo, A., & Lareu, M. V. (2013). Further development of forensic eye color predictive tests. *Forensic Science International: Genetics,* 7(1), 28–40. https://doi.org/10.1016/j.fsigen.2012.05.009

Rodrigues-Soares, F., Peñas-Lledó, E. M., Tarazona-Santos, E., Sosa-Macías, M., Terán, E., López-López, M., Rodeiro, I., Moya, G. E., Calzadilla, L. R., Ramírez-Roa, R., Grazina, M., Estévez-Carrizo, F. E., Barrantes, R., LLerena, A., Tinoco, C. A., Álvarez, M., Borbón, B., Céspedes-Garro, C., Cobaleda, J. ..., & Zamudio, R. (2020). Genomic ancestry, CYP2D6, CYP2C9, and CYP2C19 among Latin Americans. *Clinical Pharmacology and Therapeutics, 107*(1), 257–268. https://doi.org/10.1002/cpt.1598

Ruiz-Linares, A., Adhikari, K., Acuña-Alonzo, V., Quinto-Sanchez, M., Jaramillo, C., Arias, W., Fuentes, M., Pizarro, M., Everardo, P., Avila, F., Gómez-Valdés, J., León-Mimila, P., Hunemeier, T., Ramallo, V., Cerqueira, C. C. S., Burley, M. W., Konca, E., Oliveira, M. Z., Veronez, M. R. ..., & Gonzalez-José, R. (2014). Admixture in Latin America: Geographic structure, phenotypic diversity and self-perception of ancestry based on 7,342 individuals. *PLoS Genetics, 10*(9). https://doi.org/10.1371/journal.pgen.1004572

Salvoro, C., Faccinetto, C., Zucchelli, L., Porto, M., Marino, A., Occhi, G., los Campos, G., & Vazza, G. (2019). Performance of four models for eye color prediction in an Italian population sample. *Forensic Science International: Genetics, 40*(January), 192–200. https://doi.org/10.1016/j.fsigen.2019.03.008

Salzano, F. M., & Sans, M. (2014). Interethnic admixture and the evolution of Latin American populations. *Genetics and Molecular Biology, 37*(1 Suppl. 1), 151–170. https://doi.org/10.1590/S1415-47572014000200003

Samuel, G., & Prainsack, B. (2018). The regulatory landscape of forensic DNA phenotyping in Europe. *Visage, 119*. Published online https://www.visage-h2020.eu/PDF/Deliverable_5.1_for_posting_online_DECEMBER_2018.pdf.

Samuel, G., & Prainsack, B. (2019). Societal, ethical, and regulatory dimensions of forensic DNA phenotyping. *Visage, 129*. Published online https://www.visage-h2020.eu/PDF/Delliverable_5.2_for_online_publication_vo1.pdf.

Samuel, G., & Prainsack, B. (2020). Report on recommendations to address the ethical and societal challenges of FDP. *Visage, 62*. Published online https://www.visage-h2020.eu/PDF/Recommendations_for_website.pdf.

Sanchez, J. J., Phillips, C., Børsting, C., Balogh, K., Bogus, M., Fondevila, M., Harrison, C. D., Musgrave-Brown, E., Salas, A., Syndercombe-Court, D., Schneider, P. M., Carracedo, A., & Morling, N. (2006). A multiplex assay with 52 single nucleotide polymorphisms for human identification. *Electrophoresis, 27*(9), 1713−1724. https://doi.org/10.1002/elps.200500671

Serre, C., Busuttil, V., & Botto, J. M. (2018). Intrinsic and extrinsic regulation of human skin melanogenesis and pigmentation. *International Journal of Cosmetic Science, 40*(4), 328−347. https://doi.org/10.1111/ics.12466

Simcoe, M., Valdes, A., Liu, F., Furlotte, N. A., Evans, D. M., Hemani, G., Ring, S. M., Smith, G. D., Duffy, D. L., Zhu, G., Gordon, S. D., Medland, S. E., Vuckovic, D., Girotto, G., Sala, C., Catamo, E., Concas, M. P., Brumat, M., Gasparini, P. ..., & Hysi, P. G. (2021). Genome-wide association study in almost 195,000 individuals identifies 50 previously unidentified genetic loci for eye color. *Science Advances, 7*(11), 1−12. https://doi.org/10.1126/sciadv.abd1239

Slominski, A., Tobin, D. J., Shibahara, S., & Wortsman, J. (2004). Melanin pigmentation in mammalian skin and its hormonal regulation. *Physiological Reviews, 84*(4), 1155−1228. https://doi.org/10.1152/physrev.00044.2003

Söchtig, J., Phillips, C., Maroñas, O., Gómez-Tato, A., Cruz, R., Alvarez-Dios, J., Casares de Cal, M. A., Ruiz, Y., Reich, K., Fondevila, M., Carracedo, A., & Lareu, M. V. (2015). Exploration of SNP variants affecting hair colour prediction in Europeans. *International Journal of Legal Medicine, 129*(5), 963−975. https://doi.org/10.1007/s00414-015-1226-y

Solomon, O., MacIsaac, J., Quach, H., Tindula, G., Kobor, M. S., Huen, K., Meaney, M. J., Eskenazi, B., Barcellos, L. F., & Holland, N. (2018). Comparison of DNA methylation measured by Illumina 450K and EPIC BeadChips in blood of newborns and 14-year-old children. *Epigenetics, 13*(6), 655−664. https://doi.org/10.1080/15592294.2018.1497386

Soto, J., Rodriguez-Antolin, C., Vallespín, E., De Castro Carpeño, J., & Ibanez De Caceres, I. (2016). The impact of next-generation sequencing on the DNA methylation-based translational cancer research. *Translational Research, 169*, 1−18.e1. https://doi.org/10.1016/j.trsl.2015.11.003

Sulem, P., Gudbjartsson, D. F., Stacey, S. N., Helgason, A., Rafnar, T., Jakobsdottir, M., Steinberg, S., Gudjonsson, S. A., Palsson, A., Thorleifsson, G., Pálsson, S., Sigurgeirsson, B., Thorisdottir, K., Ragnarsson, R., Benediktsdottir, K. R., Aben, K. K., Vermeulen, S. H., Goldstein, A. M., Tucker, M. A. ..., & Stefansson, K. (2008). Two newly identified genetic determinants of pigmentation in Europeans. *Nature Genetics, 40*(7), 835−837. https://doi.org/10.1038/ng.160

Sulem, P., Gudbjartsson, D. F., Stacey, S. N., Helgason, A., Rafnar, T., Magnusson, K. P., Manolescu, A., Karason, A., Palsson, A., Thorleifsson, G., Jakobsdottir, M., Steinberg, S., Pálsson, S., Jonasson, F., Sigurgeirsson, B., Thorisdottir, K., Ragnarsson, R., Benediktsdottir, K. R., Aben, K. K. ..., & Stefansson, K. (2007). Genetic determinants of hair, eye and skin pigmentation in Europeans. *Nature Genetics, 39*(12), 1443−1452. https://doi.org/10.1038/ng.2007.13

The 1000 Genomes Consortium. (2015). A global reference for human genetic variation. *Nature, 526*(7571), 68−74. https://doi.org/10.1038/nature15393

Toom, V., Wienroth, M., M'Charek, A., Prainsack, B., Williams, R., Duster, T., Heinemann, T., Kruse, C., MacHado, H., & Murphy, E. (2016). Approaching ethical, legal and social issues of emerging forensic DNA phenotyping (FDP) technologies comprehensively: Reply to "Forensic DNA phenotyping: Predicting human appearance from crime scene material for investigative purposes" by Manfred Kayser. *Forensic Science International: Genetics, 22*, e1−e4. https://doi.org/10.1016/j.fsigen.2016.01.010

van den Oord, E. J. C. G., Bukszar, J., Rudolf, G., Nerella, S., McClay, J. L., Xie, L. Y., & Aberg, K. A. (2013). Estimation of CpG coverage in whole methylome next-generation sequencing studies. *BMC Bioinformatics, 14*(1). https://doi.org/10.1186/1471-2105-14-50

Vidaki, A., Ballard, D., Aliferi, A., Miller, T. H., Barron, L. P., & Syndercombe Court, D. (2017). DNA methylation-based forensic age prediction using artificial neural networks and next generation sequencing. *Forensic Science International: Genetics, 28*, 225–236. https://doi.org/10.1016/j.fsigen.2017.02.009

Walsh, S., Chaitanya, L., Breslin, K., Muralidharan, C., Bronikowska, A., Pospiech, E., Koller, J., Kovatsi, L., Wollstein, A., Branicki, W., Liu, F., & Kayser, M. (2017). Global skin colour prediction from DNA. *Human Genetics, 136*(7), 847–863. https://doi.org/10.1007/s00439-017-1808-5

Walsh, S., Chaitanya, L., Clarisse, L., Wirken, L., Draus-Barini, J., Kovatsi, L., Maeda, H., Ishikawa, T., Sijen, T., de Knijff, P., Branicki, W., Liu, F., & Kayser, M. (2014). Developmental validation of the HIrisPlex system: DNA-based eye and hair colour prediction for forensic and anthropological usage. *Forensic Science International: Genetics, 9*(1), 150–161. https://doi.org/10.1016/j.fsigen.2013.12.006

Walsh, S., Lindenbergh, A., Zuniga, S. B., Sijen, T., de Knijff, P., Kayser, M., & Ballantyne, K. N. (2011). Developmental validation of the IrisPlex system: Determination of blue and brown iris colour for forensic intelligence. *Forensic Science International: Genetics, 5*(5), 464–471. https://doi.org/10.1016/j.fsigen.2010.09.008

Walsh, S., Liu, F., Ballantyne, K. N., Van Oven, M., Lao, O., & Kayser, M. (2011). IrisPlex: A sensitive DNA tool for accurate prediction of blue and brown eye colour in the absence of ancestry information. *Forensic Science International: Genetics, 5*(3), 170–180. https://doi.org/10.1016/j.fsigen.2010.02.004

Walsh, S., Liu, F., Wollstein, A., Kovatsi, L., Ralf, A., Kosiniak-Kamysz, A., Branicki, W., & Kayser, M. (2013). The HIrisPlex system for simultaneous prediction of hair and eye colour from DNA. *Forensic Science International: Genetics, 7*(1), 98–115. https://doi.org/10.1016/j.fsigen.2012.07.005

Wendt, F. R., & Novroski, N. M. (2019). Identity informative SNP associations in the UK Biobank. *Forensic Science International: Genetics, 42*(January), 45–48. https://doi.org/10.1016/j.fsigen.2019.06.007

Wood, A. R., Esko, T., Yang, J., Vedantam, S., Pers, T. H., Gustafsson, S., Chu, A. Y., Estrada, K., Luan, J., Kutalik, Z., Amin, N., Buchkovich, M. L., Croteau-Chonka, D. C., Day, F. R., Duan, Y., Fall, T., Fehrmann, R., Ferreira, T., Jackson, A. U. ..., & Frayling, T. M. (2014). Defining the role of common variation in the genomic and biological architecture of adult human height. *Nature Genetics, 46*(11), 1173–1186. https://doi.org/10.1038/ng.3097

Woźniak, A., Heidegger, A., Piniewska-Róg, D., Pośpiech, E., Xavier, C., Pisarek, A., Kartasińska, E., Boroń, M., Freire-Aradas, A., Wojtas, M., de la Puente, M., Niederstätter, H., Płoski, R., Spólnicka, M., Kayser, M., Phillips, C., Parson, W., & Branicki, W. (2021). Development of the VISAGE enhanced tool and statistical models for epigenetic age estimation in blood, buccal cells and bones. *Aging (Albany NY), 13*(5), 6459–6484. https://doi.org/10.18632/aging.202783

Wreczycka, K., Gosdschan, A., Yusuf, D., Grüning, B., Assenov, Y., & Akalin, A. (2017). Strategies for analyzing bisulfite sequencing data. *Journal of Biotechnology, 261*(February), 105–115. https://doi.org/10.1016/j.jbiotec.2017.08.007

Xavier, C., de la Puente, M., Mosquera-Miguel, A., Freire-Aradas, A., Kalamara, V., Vidaki, A., Gross, T. E., Revoir, A., Pośpiech, E., Kartasińska, E., Spólnicka, M., Branicki, W., Ames, C. E., Schneider, P. M., Hohoff, C., Kayser, M., Phillips, C., & Parson, W. (2020). Development and validation of the VISAGE AmpliSeq basic tool to predict appearance and ancestry from DNA. *Forensic Science International: Genetics, 48*(May), 102336. https://doi.org/10.1016/j.fsigen.2020.102336

Xavier, C., de la Puente, M., Mosquera-Miguel, A., Freire-Aradas, A., Kalamara, V., Ralf, C., Revoir, A., Gross, T. E., Schneider, P. M., Ames, C., Hohoff, C., Phillips, C., Kayser, M., & Parson, W. (2022). Development and inter-laboratory evaluation of the VISAGE Enhanced Tool for Appearance and Ancestry inference from DNA. *Forensic Science International: Genetics, 61*, 102779. https://doi.org/10.1016/j.fsigen.2022.102779

Xiao, F. H., Kong, Q. P., Perry, B., & He, Y. H. (2016). Progress on the role of DNA methylation in aging and longevity. *Brief Functional Genomics, 15*(6), 454–459. https://doi.org/10.1093/bfgp/elw009

Xiong, Z., Dankova, G., Howe, L. J., Lee, M. K., Hysi, P. G., de Jong, M. A., Zhu, G., Adhikari, K., Li, D., Li, Y., Pan, B., Feingold, E., Marazita, M. L., Shaffer, J. R., McAloney, K., Xu, S., Jin, L., Wang, S., de Vrij, F. M. ..., & Kayser, M. (2019). Novel genetic loci affecting facial shape variation in humans. *bioRxiv*, 1–32. https://doi.org/10.1101/693002. Published online.

Yengo, L., Sidorenko, J., Kemper, K. E., Zheng, Z., Wood, A. R., Weedon, M. N., Frayling, T. M., Hirschhorn, J., Yang, J., & Visscher, P. M. (2018). Meta-analysis of genome-wide association studies for height and body mass index in ~700 000 individuals of European ancestry. *Human Molecular Genetics, 27*(20), 3641–3649. https://doi.org/10.1093/hmg/ddy271

Yong, W. S., Hsu, F. M., & Chen, P. Y. (2016). Profiling genome-wide DNA methylation. *Epigenetics and Chromatin, 9*(1), 1–16. https://doi.org/10.1186/s13072-016-0075-3

Yun, L., Gu, Y., Rajeevan, H., & Kidd, K. K. (2014). Application of six IrisPlex SNPs and comparison of two eye color prediction systems in diverse Eurasia populations. *International Journal of Legal Medicine, 128*(3), 447–453. https://doi.org/10.1007/s00414-013-0953-1

Zbieć-Piekarska, R., Spólnicka, M., Kupiec, T., Parys-Proszek, A., Makowska, Z., Pałeczka, A., Kucharczyk, K., Płoski, R., & Branicki, W. (2015). Development of a forensically useful age prediction method based on DNA methylation analysis. *Forensic Science International: Genetics, 17*, 173–179. https://doi.org/10.1016/j.fsigen.2015.05.001

Zubakov, D., Liu, F., Kokmeijer, I., Choi, Y., van Meurs, J. B. J., van IJcken, W. F. J., Uitterlinden, A. G., Hofman, A., Broer, L., van Duijn, C. M., Lewin, J., & Kayser, M. (2016). Human age estimation from blood using mRNA, DNA methylation, DNA rearrangement, and telomere length. *Forensic Science International: Genetics, 24*, 33–43. https://doi.org/10.1016/j.fsigen.2016.05.014

CHAPTER 17

Forensic applications of epigenetic (DNA methylation) markers through NGS

Meenu Ghai
Department of Genetics, School of Life Sciences, University of KwaZulu Natal, Durban, KwaZulu-Natal, South Africa

Introduction

Environmental factors such as lifestyle, diet, age, and stress have a significant effect on the human phenotype, which may not present itself as a change in the genome but is expressed as a modification to the epigenome (Boyce et al., 2020; Park et al., 2012; Tammen et al., 2013). The epigenome is defined as the set of chemical modifications to the DNA and DNA-associated proteins in the cell that alter gene expression and are heritable (https://www.genome.gov/about-genomics/fact-sheets/Epigenomics-Fact-Sheet). Regulation of gene expression is facilitated by differential epigenetic profiles displayed in different tissues and organs (Anastasiadi et al., 2018). Epigenetic changes include DNA methylation, nucleosomal remodeling, histone modification, chromatin looping, and noncoding RNAs. DNA methylation is the best-characterized epigenetic mark (Aristizabal et al., 2020; Goldberg et al., 2007; Mazzone et al., 2019).

DNA methylation involves the covalent attachment of the methyl group to the fifth carbon of the cytosine residue at CpG dinucleotides by the action of DNA methyltransferases. However, nonCpG DNA methylation has also been reported (Jang et al., 2017; Ramsahoye et al., 2000). Identification of differentially methylated regions (DMRs) such as tissue-specific (tDMRs), age-specific DMRs (Rakyan et al., 2011), population-specific DMRs (Hernando-Herraez et al., 2015; Heyn et al., 2013)' and differentially methylated sites (DMSs) has fascinated the field of forensics with its potential applications. For almost a decade, DMRs and DMSs have been used as candidates for biomarkers in forensic research (Kader, Ghai, & Olaniran, 2020; Kader & Ghai, 2015).

Several DNA methylation-based assays have been developed for monozygotic twin differentiation (Romanos & Borjac, 2021; Vidaki, Díez López, et al., 2017), body fluid identification (Forat et al., 2016; Ghai et al., 2020; Lee et al., 2016), and age estimation (Correia Dias et al., 2020; Weidner et al., 2014). As additional epigenetic methods are developed and validated, the need for parallel high-throughput analysis becomes important. The use of massively parallel sequencing (MPS) is well-established in clinical epigenetics and is emerging as a new technology in the forensic field (Alvarez-Cubero et al.,

Next Generation Sequencing (NGS) Technology in DNA Analysis
ISBN 978-0-323-99144-5, https://doi.org/10.1016/B978-0-323-99144-5.00013-5
© 2024 Elsevier Inc.
All rights reserved.

2017; Ballard et al., 2020). MPS, also known as next-generation sequencing (NGS), is an alternative technology to conventional capillary electrophoresis (CE), as many forensic markers can be analyzed to the sequence level in a single assay (Tytgat et al., 2022). MPS offers the ability to significantly increase the discriminatory power of human identification as well as aid in mixture deconvolution (Bennett et al., 2019).

MPS has been successfully employed for forensic SNP and STR sequencing, missing person identification, disaster victim identification, and mitochondrial DNA analysis. See Ballard et al. (2020) for a complete review of MPS in forensics. The MiSeq FGx Sequencing System is the first and only NGS instrument developed and validated for forensic genomics (Jäger et al., 2017). Though NGS techniques like pyrosequencing and Illumina have been developed for epigenetic markers, no such method has been validated for real-world case work yet.

The present chapter details the progress made regarding the utilization of NGS methods for epigenetic markers in forensics and the way forward for their routine implementation in accredited forensic laboratories. The focus is on DNA methylation markers for body fluid identification and age estimation.

Emerging MPS methods for forensics

The MPS platforms currently used for forensic applications, which can integrate epigenetic markers, are briefed below.

Illumina NovaSeq

To solve a cold case from 1976, the headless body of a pregnant woman was exhumed in 2007. Attempts were made using dental records, facial reconstruction, and fingerprinting to obtain critical information about the victim's identity. Finally, in November 2020, the DNA extract from the body was sent to Othram Inc. (https://othram.com/) for analysis. Illumina's NovaSeq was employed to construct a genealogical profile of the victim. Incidentally, the profile partially matched a DNA profile submitted by the victim's nephew to a DNA database. As a result, information about the victim was obtained from the nephew, which resulted in the arrest of the father of the victim's unborn child.

Othram Inc. (Othram.com) is the first private laboratory that utilized MPS to solve forensic cases. It has been credited for working with samples that have been regarded as unsuitable for analysis by conventional forensic technologies. There have been several reports of successful crime-solving by Othram Inc., which can be found at: https://www.solvedmysteries.com/

Illumina's MiSeq FGx sequencing system

The MiSeq FGx sequencing system allows the preparation and sequencing of libraries along with data analysis on one platform (Verogen.com, 2021). The MiSeq FGx

System is capable of analyzing degraded DNA, low-quantity DNA, complex DNA mixtures, and other challenging samples commonly encountered at a crime scene. Additionally, one sequencing run interrogates hundreds of forensically relevant genetic markers, circumventing the need to choose between fragment-length-based STR kits.

Oxford Nanopore Technologies MinION device

The Oxford Nanopore Technologies (ONT) MinION device is the newest and smallest NGS platform available, which has been used for STR and SNP profiling. One drawback of nanopore sequencing is its relatively high error rate. However, a recent study has reported accurate autosomal STR profiles from long-read sequencing data using a streamlined method called STRspy, a tool that is capable of producing length and sequence-based STR allele designations from noisy, error-prone third-generation sequencing reads. Additionally, the method allowed the detection of SNPs in the flanking regions with >90% accuracy (Hall et al., 2022). The MinION device can enable on-site analysis, which could be a boon for sensitive cases and disaster victim identification (Tytgat et al., 2022).

Ion Torrent PGM and Ion S5

The Precision ID system coupled with the Ion Torrent MPS workflow has been successfully implemented for mitochondrial DNA (mtDNA) sequencing analysis by the Missing Persons DNA Program (Cuenca et al., 2020). Ion Chef is significantly superior to Sanger's sequencing workflow as it saves time and is user-friendly. Additionally, the Precision ID MPS workflow is capable of sequencing challenging samples, particularly degraded ones that are often encountered in the Missing Persons DNA Program.

However, for nonSTR and SNP markers, the Ion Torrent MPS platform still requires further improvement. In a collaborative exercise within 17 EUROFORGEN and EDNAP laboratories, the use of MiSeq/FGx and Ion Torrent PGM/S5 platforms was optimized for the identification of body fluids using mRNA markers. The PGM/S5 workflow seemed to be less reliable with low input samples as compared to the MiSeq FGx, especially for blood, semen, or skin samples. Additionally, the PGM/S5 workflow (except for protocols including the Ion Chef robot) was reported to be more time-consuming and require more hands-on experience (e.g., chip loading) than the MiSeq FGx workflow (Ingold et al., 2018).

The present chapter focuses on DNA methylation-based MPS assays developed for body fluid identification and age estimation. Though noncoding RNA (miRNAs and Piwi RNAs) markers have also been identified for forensic applications, they are discussed in detail in another chapter of this book. A detailed review of noncoding RNAs in forensics is published by Haas et al. (2021).

Body fluid identification by MPS

The tissue specificity of DNA methylation has been identified as an attractive target for forensic applications. The benefit of DNA methylation-based methods is that the DNA isolated for STR profiling can be used for epigenetic analysis as well, although low-input protocols might be challenging to work with as bisulfite conversion degrades DNA (Peat & Smallwood, 2018) and forensic samples are also mostly poor in quality and quantity.

Pyrosequencing, being the first NGS technology developed, has been extensively used for forensic body fluid/tissue identification by DNA methylation analysis. Though clone-based bisulfite-sequencing efficiently detects methylation differences in single alleles (Huang et al., 2013), bisulfite pyrosequencing quantifies methylation levels and reliably detects differential methylation in cell populations without using the lengthy process of cloning bisulfite-treated DNA into bacterial expression vectors (Reed et al., 2010; Vidaki et al., 2016). However, to the best of my knowledge, to date, none of the pyrosequencing methods have been validated for use by accredited forensic laboratories.

The first study that employed pyrosequencing for body fluid/tissue identification was conducted by Madi et al. (2012) Gene-specific tDMRs were identified for blood, semen, and saliva. The target CpG sites at the *FGF7* gene (8 CpG sites) displayed semen-specific hypermethylation, while 5 CpG sites at *ZC3H12D* showed semen-specific hypomethylation. The C20orf117 locus differentiated blood from sperm, saliva, and epithelial cells, and the BCAS4 loci were identified as a saliva hypermethylation marker (Table 17.1).

In another study by the same research group, additional body tissues and loci were incorporated to identify spermatozoa-specific methylation markers by pyrosequencing. Blood, saliva, semen, and epithelial cells were used by Madi et al. (2012), whereas buccal cells, skin epidermis, and vaginal epithelial cells were added on by Balamurugan et al. (2014). Tissue-specific differential methylation was analyzed at the following loci: *DACT1*, *USP49*, *DDX4*, Hs_*INSL6*_03, Hs_*ZC3H12D*_05, and B_*SPTB*_03 by pyrosequencing. The markers displayed hypomethylation in semen and hypermethylation in blood, buccal cells, skin epidermis, and vaginal epithelial cells, except for one marker, B_SPTB_03, which displayed hypermethylation in semen as compared to the other tissues. The remaining three markers, C20orf117, BCAS4, and ZC3H12D, produced the same results as Madi et al. (2012). However, FGF7 did not reproduce previous results (Table 17.1).

Though pyrosequencing has been the method of choice for methylation-based body fluid identification, Bartling et al. (2014) applied Illumina MiSeq to identify semen, saliva, skin epidermis, and blood pooled in a multiplex assay. A set of eight loci, namely, L91762, L68346, L50468, L14432, L4648, L62086, L76138, and L36599, were selected from the targets used by Frumkin et al. (2011) in a previous study. A k-NN algorithm was employed to assign samples to the tissue source groups based on the normalized methylation levels. The study marked the promotion of epigenetic identification of body fluids

Table 17.1 Summary of studies reported for forensically relevant body fluid identification using MPS methods as discussed in the text.

NGS method	Sample size/tissues/body fluids included in the study	Markers identified for tissue/fluid	References
Pyrosequencing	n = 10 (blood), n = 11 (Saliva), n = 10 (sperm), n = 11 (epithelial cells)	Blood, saliva, and sperm	Madi et al. (2012)
Pyrosequencing	n = 10−11 (blood), n = 10−11 (buccal cells), n = 11−21 (skin epidermal cells), n = 9−11 (semen), n = 10 (vaginal epithelial cells)	Semen	Balamurugan et al. (2014)
Illumina Miseq	n = 4 each of saliva, venous blood, skin, and semen.	Semen, saliva, skin epidermis, and blood	Bartling et al. (2014)
Pyrosequencing	n = 100 consisting of whole blood, saliva, buccal cells, seminal fluid, vaginal fluid, menstrual secretion, skin, and urine.	Semen	Vidaki et al. (2016)
Pyrosequencing	n = 8−12 per body fluid: Blood, buccal, vaginal swabs, and semen	Vaginal epithelia	Antunes et al. (2016)
Pyrosequencing	n = 80, 20 each of blood, saliva, and vaginal secretions	Blood, saliva, semen, and vaginal secretions	Park et al. (2014)
Pyrosequencing	n = 23 (venous blood), n = 24 (buccal swabs), n = 22 (vaginal secretions), n = 20 (semen containing sperm), n = 5 (saliva), n = 3 (skin/sweat), n = 3 (semen without sperm), n = 3 (menstrual blood)	Semen, saliva, and blood	Alghanim et al. (2020)

from the conventional CE-based method to second-generation NGS. Additionally, this method could be combined with STR typing using NGS. In a similar study, Gauthier et al. (2019) implemented agglomerative hierarchical cluster analysis to create a model to indicate sample origin. A total of 5 CpG sites were selected at the loci BCAS4 (Madi et al., 2012), cg06379435 (Park et al., 2014; Silva et al., 2016), VE_8, and ZC3H12D (Madi et al., 2012) to identify saliva, blood, vaginal epithelia, and semen, respectively.

Instead of relying on previously reported differentially methylated loci, most of which were in the genic region, Park et al. (2014) identified novel CpG sites by using the Illumina Human Methylation 450 K bead array, which includes over 450,000 CpG sites. From a total of 2986 body fluid-specific tDMSs, a final eight tDMSs were selected and evaluated by pyrosequencing on DNA samples from blood, saliva, and vaginal fluid. The public dataset collected from GEO was used to obtain information for semen markers. The eight candidate tDMSs were strongly hypermethylated in the target body fluid compared to the other body fluids, which is always preferable to the markers with relative differential methylation. The following CpG sites were identified as body fluid-specific: (cg23521140 and cg17610929): semen-specific; (cg06379435 and cg08792630): blood-specific; (cg26107890 and cg20691722): saliva-specific; and (cg0177489 and cg14991487): vaginal fluid-specific. In addition to the single CpG sites discovered, the authors also analyzed neighboring CpG sites, which too displayed a hypermethylation profile similar to the target CpG site. The blood-specific CpG sites associated with age were removed to eliminate age-based bias in the identification of blood.

A combination of pyrosequencing and quantitative PCR/high-resolution melt assays was developed by Alghanim et al. (2020) for the identification of semen with sperm, semen without sperm, buccal swabs, saliva (oral fluids), venous blood, menstrual blood, vaginal secretions, and sweat/skin samples. Two novel semen-specific markers (NMUR2 and UBE2U) were shown to be effective in discriminating seminal stains containing sperm from other tissues, whereas one saliva (SA-6) and one blood (AHRR) marker were useful in identifying saliva and blood, respectively.

To identify vaginal epithelium, Antunes et al. (2016) identified a tDMR (9 CpG sites) in gene *PFN3*. Even though a significant difference in methylation was observed in vaginal epithelia when compared to blood, saliva, and semen, no clear hypo or hypermethylation was reported.

To include epigenetic markers in routine forensic work, testing their sensitivity and specificity is an utmost requirement. This necessitates proper developmental validation, similar to what is done for every STR marker or kit. To this effect, Silva et al. (2016) conducted developmental validation for three methylation-based semen (*ZC3H12D*), saliva (*BCAS4*), and blood (cg06379435)-specific markers. The study revealed that markers were primate-specific, and the sensitivity of detection varied from 0.1 to

10 ng of the starting DNA concentration. Additionally, the markers provided a consistent methylation profile for degraded and inhibited samples. Mixed body fluid samples showed varying methylation levels, and the presence of mixtures could be detected. However, complete mixture deconvolution was not achieved. Similarly, validation of two semen-specific markers (cg04382920) and (cg11768416) was carried out by Vidaki et al. (2016) to test the sensitivity and efficacy of the markers to identify aged semen samples. The pyrosequencing assay could successfully identify semen in both fresh and old semen samples (Table 17.1).

A number of DNA methylation markers have now been identified for forensically relevant body fluid identification. Hence, the next step should be to amalgamate all the markers and use MPS in a defined and validated workflow. An extensive collaborative exercise should be undertaken to assess the reproducibility of the results of methylation markers in different laboratory settings before they can be incorporated into the routine forensic workflow. Once such an exercise was performed by Jung et al. (2016) wherein seven laboratories participated to conduct a multiplex methylation SNaPshot reaction composed of seven CpG sites as markers for the identification of four body fluids, including blood, saliva, semen, and vaginal fluid. Each laboratory used four different kits for DNA extraction, five different quantification methods, two different kits for bisulfite conversion, two different PCR buffers for multiplex PCR, three different thermocyclers, three different genetic analyzers, three different analytical software, and five different analytical settings.

The performance and comparison between laboratories was assessed in four different steps, namely: (1) CE of a purified single-base extension reaction product; (2) multiplex PCR of bisulfite-modified DNA; (3) bisulfite conversion of genomic DNA; and (4) extraction of genomic DNA from body fluid samples. The results revealed that 6 out of 7 laboratories produced consistent results even though variations due to types of kits and instruments existed. In a recent follow-up study by Lee et al. (2021), a technical validation was conducted where a methylation SNaPshot assay was employed for the identification of body fluids and estimation of age at 12 different laboratories. Regarding body fluid identification, in spite of differences in the amount of bisulfite DNA used and the threshold level for a positive methylation signal, comparable results were obtained by all of the 12 laboratories, and the correct body fluid (semen, blood, and saliva) was successfully identified. The authors noted that the input volume of bisulfite-converted DNA used for PCR amplification appeared to affect the success of the analysis, and the PCR buffer composition affected the measurement of DNA methylation.

A similar exercise was undertaken by Ingold et al. (2018), wherein a total of 33 mRNA body fluid markers for blood (6), semen (6), saliva (6), vaginal fluid (4), menstrual blood (5), and skin (6) were employed and tested on the MiSeq platform, and 29 markers for blood (4), semen (6), saliva (6), vaginal fluid (4), menstrual blood (4), and

skin (5) were analyzed by the Ion Torrent platform. A total of 17 laboratories participated in the exercise; 10 used the MiSeq/FGx platform and 9 used the PGM/S5 platform (2 laboratories did the experiments on both platforms). Furthermore, epigenetic markers should also be tested in combination with other genotyping markers on an MPS platform to assess how they complement each other. Such a collaborative exercise was conducted by Ingold et al. (2020), wherein 35 RNA-cSNPs and 33 mRNA markers were tested in two different multiplex assays on MiSeq/FGx platforms in nine laboratories.

Age estimation by MPS

DNA methylation patterns change with increasing age and contribute to age-related diseases (Bocklandt et al., 2011). Epigenetic clocks use predictable changes in the epigenome, usually DNA methylation, to estimate chronological age with unprecedented accuracy (Ryan, 2021). Several articles have reported age predictors based on DNA methylation levels in specific tissues, for example, saliva or blood (Bocklandt et al., 2011; Hannum et al., 2013); however, research on age estimation using DNA methylation attracted the forensic world since 2013, with the report of the epigenetic clock, which could be used as a multi-tissue age predictor by Horvath (2013). He identified and characterized a total of 353 CpG sites that together form an aging clock.

To develop an epigenetic age prediction model, two sample groups are required: a training group and an evaluation group. A training group is used for the initial building of the model, and an evaluation or validation group is used to evaluate the prediction's performance. Both pyrosequencing and Illumina MiSeq have taken center stage for forensic age estimation. The studies below detail the evolution of DNA methylation-based age prediction.

Before NGS was used for age prediction, Garagnani et al. (2012) employed an Illumina Infinium HumanMethylation450 bead chip to identify age-associated CpG sites in genes *ELOVL2*, *FHL2*, and *PENK* from the whole blood of 32 mother-offspring couples aged 42–83 and 9–52, respectively. The CpG islands of *ELOVL2*, *FHL2*, and *PENK* in the gene promoters were found to be hypermethylated in mothers compared to the offspring. The results were also replicated in 494 individuals (245 men and 249 women) ranging in age from 9 to 99 years by Sequenom's EpiTYPER assay.

Later, Weidner et al. (2014) identified three age-related CpGs located in the genes *ITGA2B*, *ASPA,* and *PDE4C* by bisulfite pyrosequencing of 151 blood samples that could robustly predict age with a mean absolute deviation (MAD) from chronological age of fewer than 5 years. This approach requires locus-specific pyrosequencing and is cost-effective. The study also revealed that patients with acquired aplastic anemia or dyskeratosis congenita were predicted to be prematurely aged due to progressive bone marrow failure and severe telomere attrition (Table 17.2).

Table 17.2 Summary of studies reported for age prediction by MPS methods as discussed in the text.

NGS method	Sample size/training set (T) + validation set (V) for age prediction model development[a]	Target tissue/body fluid	Age range	References
Pyrosequencing	*n = 82 (T) + 69 (V)*	Blood	0–78 years	Weidner et al. (2014)
Pyrosequencing	n = 44 (saliva), n = 23 (blood)	Saliva, blood	5–72 years	Soares Bispo Santos Silva et al. (2015)
Pyrosequencing	n = 169 deceased, n = 37 living. n = 29 healthy erupted third molars	Blood	0–91 years, dental samples: 19–70 years	Bekaert et al. (2015b)
Pyrosequencing	*303 (T) + 124 (V)*	Blood	2–75 years	Zbieć-Piekarska, Spólnicka, Kupiec, Makowska, et al. (2015)
Pyrosequencing	*300 (T) + 120 (V)*	Blood	2–75 years	Zbieć-Piekarska, Spólnicka, Kupiec, Parys-Proszek, et al. (2015)
Pyrosequencing	*535 (T) + 230 (V)*	Blood	11–90 years	Park et al. (2016)
Pyrosequencing	Saliva: *52 (T) + 39(V)*, blood: 40 (T) + 32 (V)	Blood, saliva	9–73 years, blood: 5–72 years	Alghanim et al. (2017)
Illumina MiSeq	n = 1156 (blood) n = 265 (saliva) *159 (T) + 53 (V) + 53 (blind test)* n = 106, blood samples from 53 monozygotic twin pairs n = 1011, blood samples, (577 females and 434 males) from diseased donors	Blood, saliva	Blood: 2–90 years Saliva- 21–55 years Monozygotic twins: 33–77 years Diseased samples: 17–91 years	Vidaki, Ballard, et al., (2017)

Continued

Table 17.2 Summary of studies reported for age prediction by MPS methods as discussed in the text.—cont'd

NGS method	Sample size/training set (T) + validation set (V) for age prediction model development[a]	Target tissue/body fluid	Age range	References
Pyrosequencing	*100 (T) + 36 (V)*	Postmortem blood	18–60 years	Sukawutthiya et al. (2021)
Pyrosequencing and illumina MiSeq	84	Blood	18–99 years	Freire-Aradas et al. (2020)
MiSeq	144 (29 each tissue/fluid) except n = 28 (buccal swabs)	Samples of brain, bone, muscle, buccal swabs, and whole blood	0–87 years	Naue et al. (2018)
MiSeq/FGx	31	Plucked hair	5–88 years	Naue et al. (2021)
MiSeq/FGx	n = 160 (blood), n = 160 (buccal swab), n = 161 (bone) *112 (T) + 48 (V)* (blood and buccal cells) + *49 (V)* (bones)	Blood, buccal cells, bone	Blood: 1–75 years, Buccal cells: 2–80 years, bone: 19–93 years.	Woźniak et al. (2021)
MiSeq	90 exogenous deaths *60 (T) + 30 (V)*	Postmortem blood	2-weeks–91 years	Guan et al. (2021)
Pyrosequencing	n = 73 deceased 68.5% male) n = 142 healthy living 41.5% male).	Buccal swabs in different stages of decomposition	Deceased: 0–90 years Healthy: 0–89 years	Koop et al. (2021)

[a]Number of samples used for age prediction model development are indicated in italics.

Age-correlated hypermethylated CpG sites were reported by Soares Bispo Santos Silva et al. (2015) in *GRIA2* (3 CpG sites) and *NPTX2* genes (6 CpG sites). The MAD was 6.9 and 9.2 for the *GRIA2* and *NPTX2* loci, respectively. Older individuals displayed increased methylation in both markers compared to younger individuals, and this trend was more pronounced in the GRIA2 locus when compared to NPTX2.

As pyrosequencing-based age determination assays developed, both living and deceased blood samples were utilized by Bekaert et al. (2015a), and results demonstrated that methylation of CpG sites in the *ASPA* and *EDARADD* genes was negatively correlated with age, whereas positive correlations were found between methylation and age for *PDE4C* and *ELOVL2*. The multivariate quadratic regression model predicted donor age with a mean error of 3.75 years. Additionally, the authors conducted an age estimation using dentin samples with a MAD of 4.86 years.

In addition to the frequently used gene *ELOVL2* Zbieć-Piekarska, Spólnicka, Kupiec, Parys-Proszek, et al. (2015) identified additional age-correlated CpG sites at C1orf132, TRIM59, KLF14, and FHL2 loci. With a training set of 300 individuals, a MAD of 3.4 years was observed, and 120 individuals were used to develop the validation set with a MAD of 3.9 years. The age predictor was superior for young individuals as compared to older ones. Also, the gender difference was reported in the training set, with females showing lower MAD (Table 17.2).

DNA methylation is population-specific (Galanter et al., 2017; Kader, Ghai, & Zhou, 2020). To study age prediction in the Asian population, Park et al. (2016) evaluated methylation levels at three age-correlated CpG sites, namely cg16867657 (*ELOVL2*), cg04208403 (*ZNF423*), and cg19283806 (*CCDC102B*), by pyrosequencing. A total of 535 samples were used to develop an age prediction model with a MAD of 3.156 years, and the validation set consisted of 230 samples with a MAD of 3.346. Increased accuracy was observed for the <60-year-old age group and 57.30% in the older group (≥60 years).

Meanwhile, Guan et al. (2021) were interested in examining the reliability of epigenetic age markers in postmortem blood and specifically in the Japanese population. DNA methylation analysis was done using the Illumina MiSeq system to identify age-associated CpGs. Four CpG sites on four genes—*ASPA*, *ELOVL2*, *ITGA2B*, and *PDE4C*—were found to be highly correlated with chronological age. Whole blood samples were collected from 2-week-olds to 91-year-olds during autopsies, and it was ensured that the cause of death was extrinsic (due to strangulation and lethal trauma and not from any disease) and the postmortem interval for each subject was less than 10 days to avoid severe DNA fragmentation and contamination (Table 17.2).

All epigenetic age models are generally based on regression-based prediction with a certain range of MAD. In order to increase the model accuracy and decrease error, Vidaki, Ballard, et al. (2017) utilized machine learning algorithms, specifically their artificial neural networks (ANNs), combined with NGS (Illumina MiSeq) for age

prediction. ANNs are a group of machine-learning algorithms inspired by biological systems. This is the first study that used machine learning, via ANNs, together with an NGS-based DNA methylation detection method for forensic age prediction. Both blood and saliva samples were used for the study. The age prediction model was validated by an independent cohort of 53 monozygotic twin pairs aged 33–77 years. Few between-pair twins were predicted to be either much older or much younger than their actual age, but the prediction differences within twin pairs were not statistically significant. The authors also investigated a set of diseased samples and their effect on age prediction. The mean absolute errors for each disease differed as follows: type I diabetes—8.63 years; anemia—14.38 years; bone marrow disorders (including leukemia)—11.09 years; ovarian cancer—7.45 years; breast cancer—6.77 years; and schizophrenia—5.03 years, which could influence epigenetic age prediction.

Additional age prediction models were developed using bisulfite pyrosequencing for forensic age estimation by Alghanim et al. (2017) targeting CpG sites within genes *SCGN*, *DLX5,* and *KLF14*. Sukawutthiya et al. (2021) used only 2 CpG sites in the gene *ELOVL2* to obtain good age indicators in the age groups of 20–40 years and over 40 years. Decedent blood was used by Anaya et al. (2021) to predict chronological age by bisulfite pyrosequencing analysis of single CpG sites on five genes: *KLF14*, *ELOVL2*, *C1orf132*, *TRIM59*, and *FHL2*. It was observed that the prediction potential of the models decreased with the increasing age of the person, particularly for the age categories above 50 years.

It is vital to compare the developed age prediction models in an independent validation study using the same population to select the best one for forensic applications. Hence, Daunay et al. (2019) evaluated and compared six prediction models (Bekaert et al., 2015a; Park et al., 2016; Thong et al., 2017; Weidner et al., 2014; Zbieć-Piekarska, Spólnicka, Kupiec, Makowska, et al., 2015; Zbieć-Piekarska, Spólnicka, Kupiec, Parys-Proszek, et al., 2015) in an independent study using 100 blood samples from French individuals aged between 19 and 65 years. The Bekaert model presented the best overall performance and accuracy for age prediction (MAD of 4.5 and SEE of 6.8), followed by the Thong model (MAD of 5.2 and SEE of 7.2). However, for forensic purposes, the Zbiec-Piekarska 1 model was found to be most suitable as it allows quick and accurate age estimation even with low quantities of DNA, based on just 2 CpG sites located at a single locus.

Four different technologies: EpiTYPER, pyrosequencing, MiSeq, and SNaPshot were utilized for epigenetic age estimation of DNA from 84 blood samples from healthy Europeans ranging in age from 18 to 99 years (Freire-Aradas et al., 2020). DNA methylation was analyzed at 4 CpG sites for three genes, viz., *ELOVL2*, *FHL2*, and *MIR29B2,* which have been previously reported to be age-specific (Jung et al., 2019; Zbieć-Piekarska, Spólnicka, Kupiec, Parys-Proszek, et al., 2015). At ELOVL2 and FHL2 loci, similar

patterns of DNA methylation for EpiTYPER, pyrosequencing, and MiSeq were observed, and subsequently, data from these techniques can be used in platform-independent age prediction models.

Several reports have been documented above on age prediction models based on age-correlated CpGs. However, to enable the routine application of methylation-based age prediction, standardization and reproducibility of these methods are necessary. Therefore, Pfeifer et al. (2020) tested two different published age prediction models for blood and mouth swab samples with regards to prediction accuracy (Bekaert et al., 2015a, 2015b).

A total of 151 whole blood samples (aged 1 day−96 years) of deceased individuals and 149 mouth swabs (aged 13 days−95 years) of living individuals were collected from 63 or 60 female (blood/buccal swabs) and 88 or 89 male (blood/buccal swabs) donors. The mean time between death and sample collection was 6 days, with a minimum of less than 24 h and a maximum of 20 days. Additionally, freshly taken blood samples of 21 individuals (20−65 years old) were collected from donors with disease states that included different cancers, mental and physical handicaps, different levels of diabetes, lung diseases (chronic bronchitis and asthma), schizophrenia, and coronary heart diseases. Moreover, drug and alcohol abuse could be documented for 20 deceased individuals. Results generated revealed that the CpG site in the *ELOVL2* gene was the most informative age-associated marker, which displayed hypermethylation with increasing age. CpG sites in genes *ASPA*, *CCDC102B*, *C1orf132*, ITGA2B, *KLF14,* and *PDE4C* displayed a moderate correlation with age. The postmortem period and diseases had no effect on DNA methylation (Table 17.2).

The authors demonstrated that the interlaboratory variance of variable sampling, different DNA inputs, bisulfite conversion systems, PCR methods, and analysis methods greatly influence the quantification of DNA methylation. To mitigate the effect of the above variables, the use of standards to normalize DNA methylation results for better comparison with other studies could be helpful.

In addition to blood and saliva, MPS has been utilized to determine age using not-so-commonly used tissues like the brain, bone, muscle, buccal swabs (Naue et al., 2018; Woźniak et al., 2021), and hair (Naue et al., 2021). The hair was plucked from 49 deceased individuals, both male and female (aged 5−88). Hair was from different phases of the hair cycle, i.e., the growing phase, transition phase, and resting phase, as given at the time of the individual's death. Samples were sequenced on a MiSeq FGx (Verogen, San Diego, CA, USA). From a total of 49 samples, only 31 could be used for methylation determination, as 18 samples showed low sequencing coverage or bisulfite conversion efficiency below 98%. This could be because of the use of hair, which often contains a low quantity of DNA or fragmented DNA. A high correlation between the DNAm and age was observed for the ELOVL2, KLF14, RPA2, TRIM59, and ZYG11A loci.

In a proof-of-principle study, Koop et al. (2021) used buccal swabs under different postmortem conditions, viz., with visible signs of no decomposition, early decomposition, advanced decomposition, and severe decomposition for epigenetic age estimation by pyrosequencing at the age-associated CpG-1 site of the *PDE4C* (upstream of cg17861230) gene. Postmortem samples, except severely decomposed ones, exhibited a high correlation of methylation levels with age, independent of the state of decomposition, which was comparable to the correlation for samples from live donors.

The gradual process by which epigenetic markers could be incorporated into the current forensic workflow was envisioned by Heidegger et al. (2022) who presented the first interlaboratory validation of a single targeted bisulfite MPS assay (Illumina MiSeq) for forensic age prediction. The assay was validated by five consortium laboratories of the VISible Attributes through GEnomics (VISAGE) project. Two panels of age-specific markers were tested. One consisted of 13 markers (ET-13), and the second consisted of 5 markers (ET-5). The study thoroughly looked at the reproducibility, species specificity, and sensitivity of producing DNA methylation profiles from semen samples. Both the 13 and 5-marker sets performed well. Methylation quantification was reproducible between laboratories except for the fact that variability in methylation quantification results increased below 50 ng DNA, thus affecting the difference between technical duplicates.

Improvements in the application of epigenetic markers via MPS

Though MPS allows multiplexing and analysis of highly degraded DNA samples, it is necessary to improve the analytical tools in order to obtain more reliable results for epigenetic markers. Currently, many analysis algorithms are not standardized (Bruijns et al., 2018; Jordan & Mills, 2021). Additionally, for some target CpGs, a large interindividual variation in methylation levels has been observed (Vidaki et al., 2016), since DNA methylation patterns are dynamic and can be influenced by environmental factors. Hence, several confounding factors could affect identification (e.g., body fluids) and predictions (e.g., age estimation) and must be taken into account (Jordan & Mills, 2021).

Future considerations

MPS or NGS methods have successfully been integrated into CE-based workflows for forensic human identification. The next step is to set up a parallel platform for the application of epigenetic markers in forensic work that can complement the existing technology. The most essential prerequirement for any methylation-based marker is bisulfite conversion of DNA, which highly deteriorates DNA's quality. Therefore, methylation detection methods need to be verified and evaluated to meet the requirements of high sensitivity for low-template DNA samples. To improve reproducibility

among laboratories, the standard workflow of forensic epigenetic analyses is an important factor to be considered. Pyrosequencing for body fluid identification and age estimation has been extensively verified and validated. In the future, it would be worthwhile to multiplex markers for body fluid identification and age estimation into a single reaction. Similar to the collaborative exercise conducted by several labs for mRNA-based body fluid identification using MPS methods, forensic labs should also collaborate to test epigenetic markers across laboratories to evaluate the MPS application in an extended group.

Age-prediction models should include data on standards with known DNA methylation values for every pyrosequencing assay for normalization of DNA methylation data for a better comparison between studies. Furthermore, DNA methylation differences between ethnicities/ancestries should be identified, and markers unaffected by population differences should be developed to ensure accurate age estimation. Moreover, a systematic evaluation of the different age prediction models on the same population should be performed in order to compare their performance and identify the model with the best age prediction accuracy. Validation studies are essential to understanding technical assay variability.

NGS allows improved resolution and simultaneous analysis of STRs, SNPs, and Y- and X- chromosome markers. In the future, epigenetic methods such as the DNA methylation profile at target CpGs will complement STRs and SNPs. A new approach of CpG-SNPs for individual-specific body fluid identification was reported by Watanabe et al. (2022). Even though this approach is still in its infancy, the use of CpG-SNPs could be an effective method to indicate SNPs associated with body fluids in a mixture. Additionally, DNA methylation assays could be used to demonstrate the lifestyle of individuals, such as smoking status (Alghanim et al., 2018).

References

Alghanim, H., Antunes, J., Silva, D. S. B. S., Alho, C. S., Balamurugan, K., & McCord, B. (November 2017). Detection and evaluation of DNA methylation markers found at SCGN and KLF14 loci to estimate human age. *Forensic Science International: Genetics, 31*, 81–88. https://doi.org/10.1016/j.fsigen.2017.07.011

Alghanim, H., Balamurugan, K., & McCord, B. R. (2020). Development of DNA methylation markers for sperm, saliva and blood identification using pyrosequencing and qPCR/HRM. *Analytical Biochemistry*, 113933. https://doi.org/10.1016/j.fsigen.2016.01.017

Alghanim, H., Wu, W., & McCord, B. (November 2018). DNA methylation assay based on pyrosequencing for determination of smoking status. *Electrophoresis, 39*(21), 2806–2814. https://doi.org/10.1002/elps.201800098

Alvarez-Cubero, M. J., Saiz, M., Martínez-García, B., Sayalero, S. M., Entrala, C., Lorente, J. A., Martinez-Gonzalez, L. J., et al. (2017). Next generation sequencing: An application in forensic sciences. *Annals of Human Biology, 44*(7), 581–592. https://doi.org/10.1080/03014460.2017.1375155

Anastasiadi, D., Esteve-Codina, A., & Piferrer, F. (June 29, 2018). Consistent inverse correlation between DNA methylation of the first intron and gene expression across tissues and species. *Epigenetics & Chromatin, 11*(1), 37. https://doi.org/10.1186/s13072-018-0205-1

Anaya, Y., Yew, P., Roberts, K. A., & Hardy, W. R. (November 2021). DNA methylation of decedent blood samples to estimate the chronological age of human remains. *International Journal of Legal Medicine, 135*(6), 2163−2173. https://doi.org/10.1007/s00414-021-02650-8

Antunes, J., Silva, D. S., Balamurugan, K., Duncan, G., Alho, C. S., & McCord, B. (October 2016). Forensic discrimination of vaginal epithelia by DNA methylation analysis through pyrosequencing. *Electrophoresis, 37*(21), 2751−2758. https://doi.org/10.1002/elps.201600037

Aristizabal, M. J., Anreiter, I., Halldorsdottir, T., Odgers, C. L., McDade, T. W., Goldenberg, A., Mostafavi, S., Kobor, M. S., Binder, E. B., Sokolowski, M. B., & O'Donnell, K. J. (September 22, 2020). Biological embedding of experience: A primer on epigenetics. *Proceedings of the National Academy of Sciences of the U S A, 117*(38), 23261−23269. https://doi.org/10.1073/pnas.1820838116

Balamurugan, K., Bombardi, R., Duncan, G., & McCord, B. (November 2014). Identification of spermatozoa by tissue-specific differential DNA methylation using bisulfite modification and pyrosequencing. *Electrophoresis, 35*(21−22), 3079−3086. https://doi.org/10.1002/elps.201400175

Ballard, D., Winkler-Galicki, J., & Wesoły, J. (July 2020). Massive parallel sequencing in forensics: Advantages, issues, technicalities, and prospects. *International Journal of Legal Medicine, 134*(4), 1291−1303. https://doi.org/10.1007/s00414-020-02294-0

Bartling, C. M., Hester, M. E., Bartz, J., Heizer, E., Jr., & Faith, S. A. (November 2014). Next-generation sequencing approach to epigenetic-based tissue source attribution. *Electrophoresis, 35*(21−22), 3096−3101. https://doi.org/10.1002/elps.201400087

Bekaert, B., Kamalandua, A., Zapico, S. C., Van de Voorde, W., & Decorte, R. (2015a). Improved age determination of blood and teeth samples using a selected set of DNA methylation markers. *Epigenetics, 10*(10), 922−930. https://doi.org/10.1080/15592294.2015.1080413

Bekaert, B., Kamalandua, A., Zapico, S. C., Van de Voorde, W., & Decorte, R. (2015b). A selective set of DNA-methylation markers for age determination of blood, teeth and buccal samples. *Forensic Science International: Genetics Supplement Series*, (5), e144−e145. https://doi.org/10.1016/j.fsigss.2015.09.058

Bennett, L., Oldoni, F., Long, K., Cisana, S., Maddela, K., Wootton, S., Chang, J., Hasegawa, R., Lagacé, R., Kidd, K. K, & Podini, D. (2019). Mixture deconvolution by massively parallel sequencing of microhaplotypes [published correction appears in Int J Legal Med. 2019 Mar 5. Madella K [corrected to Maddela K]] *International Journal of Legal Medicine, 133*(3), 719−729. https://doi.org/10.1007/s00414-019-02010-7

Bocklandt, S., Lin, W., Sehl, M. E., Sánchez, F. J., Sinsheimer, J. S., Horvath, S., & Vilain, E. (2011). Epigenetic predictor of age. *PLoS One, 6*(6), e14821. https://doi.org/10.1371/journal.pone.0014821

Boyce, W. T., Sokolowski, M. B., & Robinson, G. E. (September 22, 2020). Genes and environments, development and time. *Proceedings of the National Academy of Sciences of the U S A, 117*(38), 23235−23241. https://doi.org/10.1073/pnas.2016710117

Bruijns, B., Tiggelaar, R., & Gardeniers, H. (November 2018). Massively parallel sequencing techniques for forensics: A review. *Electrophoresis, 39*(21), 2642−2654. https://doi.org/10.1002/elps.201800082

Correia Dias, H., Cunha, E., Corte Real, F., & Manco, L. (November 2020). Age prediction in living: Forensic epigenetic age estimation based on blood samples. *Legal Medicine, 47*, 101763. https://doi.org/10.1016/j.legalmed.2020.101763

Cuenca, D., Battaglia, J., Halsing, M., & Sheehan, S. (November 4, 2020). Mitochondrial sequencing of missing persons DNA casework by implementing thermo Fisher's precision ID mtDNA whole genome assay. *Genes, 11*(11), 1303. https://doi.org/10.3390/genes11111303

Daunay, A., Baudrin, L. G., Deleuze, J. F., & How-Kit, A. (June 20, 2019). Evaluation of six blood-based age prediction models using DNA methylation analysis by pyrosequencing. *Scientific Reports, 9*(1), 8862. https://doi.org/10.1038/s41598-019-45197

Forat, S., Huettel, B., Reinhardt, R., Fimmers, R., Haidl, G., Denschlag, D., & Olek, K. (February 1, 2016). Methylation markers for the identification of body fluids and tissues from forensic trace evidence. *PLoS One, 11*(2), e0147973. https://doi.org/10.1371/journal.pone.0147973. Erratum in: PLoS One. 2016; 11(5):e0156472.

Freire-Aradas, A., Pośpiech, E., Aliferi, A., Girón-Santamaría, L., Mosquera-Miguel, A., Pisarek, A., Ambroa-Conde, A., Phillips, C., Casares de Cal, M. A., Gómez-Tato, A., Spólnicka, M., Woźniak, A., Álvarez-Dios, J., Ballard, D., Court, D. S., Branicki, W., Carracedo, Á., &

Lareu, M. V. (August 19, 2020). A comparison of forensic age prediction models using data from four DNA methylation technologies. *Frontiers in Genetics, 11*, 932. https://doi.org/10.3389/fgene.2020.00932

Frumkin, D., Wasserstrom, A., Budowle, B., & Davidson, A. (November 2011). DNA methylation-based forensic tissue identification. *Forensic Science International: Genetics, 5*(5), 517–524. https://doi.org/10.1016/j.fsigen.2010.12.001

Galanter, J. M., Gignoux, C. R., Oh, S. S., Torgerson, D., Pino-Yanes, M., Thakur, N., Eng, C., Hu, D., Huntsman, S., Farber, H. J., Avila, P. C., Brigino-Buenaventura, E., LeNoir, M. A., Meade, K., Serebrisky, D., Rodríguez-Cintrón, W., Kumar, R., Rodríguez-Santana, J. R., Seibold, M. A., Borrell, L. N., ...Zaitlen, N., et al. (2017). Differential methylation between ethnic sub-groups reflects the effect of genetic ancestry and environmental exposures. *Elife, 6*, e20532. https://doi.org/10.7554/eLife.20532

Garagnani, P., Bacalini, M. G., Pirazzini, C., Gori, D., Giuliani, C., Mari, D., Di Blasio, A. M., Gentilini, D., Vitale, G., Collino, S., Rezzi, S., Castellani, G., Capri, M., Salvioli, S., & Franceschi, C. (December 2012). Methylation of ELOVL2 gene as a new epigenetic marker of age. *Aging Cell, 11*(6), 1132–1134. https://doi.org/10.1111/acel.12005

Gauthier, Q. T., Cho, S., Carmel, J. H., & McCord, B. R. (September 2019). Development of a body fluid identification multiplex via DNA methylation analysis. *Electrophoresis, 40*(18–19), 2565–2574. https://doi.org/10.1002/elps.201900118

Ghai, M., Naidoo, N., Evans, D. L., & Kader, F. (November 2020). Identification of novel semen and saliva specific methylation markers and its potential application in forensic analysis. *Forensic Science International: Genetics, 49*, 102392. https://doi.org/10.1016/j.fsigen.2020.102392

Goldberg, A. D., Allis, C. D., & Bernstein, E. (February 23, 2007). Epigenetics: A landscape takes shape. *Cell, 128*(4), 635–638. https://doi.org/10.1016/j.cell.2007.02.006

Guan, X., Ohuchi, T., Hashiyada, M., & Funayama, M. (November 2021). Age-related DNA methylation analysis for forensic age estimation using post-mortem blood samples from Japanese individuals. *Legal Medicine, 53*, 101917. https://doi.org/10.1016/j.legalmed.2021.101917

https://www.genome.gov/about-genomics/fact-sheets/Epigenomics-Fact-Sheet. www.genome.gov/. Accessed on December 22, 2021.

https://www.solvedmysteries.com/. Accessed on December 15, 2021.

Haas, C., Neubauer, J., Salzmann, A. P., Hanson, E., & Ballantyne, J. (May 2021). Forensic transcriptome analysis using massively parallel sequencing. *Forensic Science International: Genetics, 52*, 102486. https://doi.org/10.1016/j.fsigen.2021.102486

Hall, C. L., Kesharwani, R. K., Phillips, N. R., Planz, J. V., Sedlazeck, F. J., & Zascavage, R. R. (January 2022). Accurate profiling of forensic autosomal STRs using the Oxford Nanopore Technologies MinION device. *Forensic Science International: Genetics, 56*, 102629. https://doi.org/10.1016/j.fsigen.2021.102629

Hannum, G., Guinney, J., Zhao, L., Zhang, L., Hughes, G., Sadda, S., Klotzle, B., Bibikova, M., Fan, J. B., Gao, Y., Deconde, R., Chen, M., Rajapakse, I., Friend, S., Ideker, T., & Zhang, K. (2013). Genome-wide methylation profiles reveal quantitative views of human aging rates. *Molecular Cell, 49*(2), 359–367. https://doi.org/10.1016/j.molcel.2012.10.016

Heidegger, A., Pisarek, A., de la Puente, M., Niederstätter, H., Pośpiech, E., Woźniak, A., Schury, N., Unterländer, M., Sidstedt, M., Junker, K., Ventayol Garcia, M., Laurent, F. X., Ulus, A., Vannier, J., Bastisch, I., Hedman, J., Sijen, T., Branicki, W., Xavier, C., Parson, W., & VISAGE Consortium. (January 2022). Development and inter-laboratory validation of the VISAGE enhanced tool for age estimation from semen using quantitative DNA methylation analysis. *Forensic Science International: Genetics, 56*, 102596. https://doi.org/10.1016/j.fsigen.2021.102596

Hernando-Herraez, I., Garcia-Perez, R., Sharp, A. J., & Marques-Bonet, T. (December 10, 2015). DNA methylation: Insights into human evolution. *PLoS Genetics, 11*(12), e1005661. https://doi.org/10.1371/journal.pgen.1005661

Heyn, H., Moran, S., Hernando-Herraez, I., Sayols, S., Gomez, A., Sandoval, J., Monk, D., Hata, K., Marques-Bonet, T., Wang, L., & Esteller, M. (September 2013). DNA methylation contributes to natural human variation. *Genome Research, 23*(9), 1363–1372. https://doi.org/10.1101/gr.154187.112

Horvath, S. (2013). DNA methylation age of human tissues and cell types [published correction appears in Genome Biol. 2015; 16:96]. *Genome Biology, 14*(10), R115. https://doi.org/10.1186/gb-2013-14-10-r115

Huang, Z., Bassil, C. F., & Murphy, S. K. (2013). Bisulfite sequencing of cloned alleles. *Methods in Molecular Biology, 1049*, 83–94. https://doi.org/10.1007/978-1-62703-547-7_8

Ingold, S., Dørum, G., Hanson, E., Ballard, D., Berti, A., Gettings, K. B., Giangasparo, F., Kampmann, M. L., Laurent, F. X., Morling, N., Parson, W., Steffen, C. R., Ulus, A., van den Berge, M., van der Gaag, K. J., Verdoliva, V., Xavier, C., Ballantyne, J., & Haas, C. (March 2020). Body fluid identification and assignment to donors using a targeted mRNA massively parallel sequencing approach – results of a second EUROFORGEN/EDNAP collaborative exercise. *Forensic Science International: Genetics, 45*, 102208. https://doi.org/10.1016/j.fsigen.2019.102208

Ingold, S., Dørum, G., Hanson, E., Berti, A., Branicki, W., Brito, P., Elsmore, P., Gettings, K. B., Giangasparo, F., Gross, T. E., Hansen, S., Hanssen, E. N., Kampmann, M. L., Kayser, M., Laurent, F. X., Morling, N., Mosquera-Miguel, A., Parson, W., Phillips, C., & Haas, C. (May 2018). Body fluid identification using a targeted mRNA massively parallel sequencing approach – results of a EUROFORGEN/EDNAP collaborative exercise. *Forensic Science International: Genetics, 34*, 105–115. https://doi.org/10.1016/j.fsigen.2018.01.002

Jäger, A. C., Alvarez, M. L., Davis, C. P., Guzmán, E., Han, Y., Way, L., Walichiewicz, P., Silva, D., Pham, N., Caves, G., Bruand, J., Schlesinger, F., Pond, S. J. K., Varlaro, J., Stephens, K. M., & Holt, C. L. (May 2017). Developmental validation of the MiSeq FGx forensic genomics system for targeted next generation sequencing in forensic DNA casework and database laboratories. *Forensic Science International: Genetics, 28*, 52–70. https://doi.org/10.1016/j.fsigen.2017.01.011

Jang, H. S., Shin, W. J., Lee, J. E., & Do, J. T. (May 23, 2017). CpG and non-CpG methylation in epigenetic gene regulation and brain function. *Genes, 8*(6), 148. https://doi.org/10.3390/genes8060148

Jordan, D., & Mills, D. (2021). Past, present, and future of DNA typing for analyzing human and non-human forensic samples. *Frontiers in Ecology and Evolution, 9*, 646130. https://doi.org/10.3389/fevo.2021.646130

Jung, S. E., Cho, S., Antunes, J., Gomes, I., Uchimoto, M. L., Oh, Y. N., Di Giacomo, L., Schneider, P. M., Park, M. S., van der Meer, D., Williams, G., McCord, B., Ahn, H. J., Choi, D. H., Lee, Y. H., Lee, S. D., & Lee, H. Y. (October 2016). A collaborative exercise on DNA methylation-based body fluid typing. *Electrophoresis, 37*(21), 2759–2766. https://doi.org/10.1002/elps.201600256

Jung, S. E., Lim, S. M., Hong, S. R., Lee, E. H., Shin, K. J., & Lee, H. Y. (2019). DNA methylation of the ELOVL2, FHL2, KLF14, C1orf132/MIR29B2C, and TRIM59 genes for age prediction from blood, saliva, and buccal swab samples. *Forensic Science International: Genetics, 38*, 1–8. https://doi.org/10.1016/j.fsigen.2018.09.010

Justice through genomics. https://othram.com. Accessed on December 20, 2021.

Kader, F., & Ghai, M. (April 2015). DNA methylation and application in forensic sciences. *Forensic Science International, 249*, 255–265. https://doi.org/10.1016/j.forsciint.2015.01.037

Kader, F., Ghai, M., & Olaniran, A. O. (2020). Characterization of DNA methylation-based markers for human body fluid identification in forensics: A critical review. *International Journal of Legal Medicine, 134*(1), 1–20. https://doi.org/10.1007/s00414-019-02181-3

Kader, F., Ghai, M., & Zhou, M. (2020). Ethnicity, age and disease-associated variation in body fluid-specific CpG sites in a diverse South African cohort. *Forensic Science International, 314*, 110372. https://doi.org/10.1016/j.forsciint.2020.110372

Koop, B. E., Mayer, F., Gündüz, T., Blum, J., Becker, J., Schaffrath, J., Wagner, W., Han, Y., Boehme, P., & Ritz-Timme, S. (January 2021). Postmortem age estimation via DNA methylation analysis in buccal swabs from corpses in different stages of decomposition-a "proof of principle" study. *International Journal of Legal Medicine, 135*(1), 167–173. https://doi.org/10.1007/s00414-020-02360-7

Lee, H. Y., Jung, S. E., Lee, E. H., Yang, W. I., & Shin, K. J. (September 2016). DNA methylation profiling for a confirmatory test for blood, saliva, semen, vaginal fluid and menstrual blood. *Forensic Science International: Genetics, 24*, 75–82. https://doi.org/10.1016/j.fsigen.2016.06.007

Lee, J. E., Lee, J. M., Naue, J., Fleckhaus, J., Freire-Aradas, A., Neubauer, J., Pośpiech, E., McCord, B., Kalamara, V., Gauthier, Q., Mills, C., Cao, Y., Wang, Z., Oh, Y. N., Feng, L., Schneider, P. M.,

Phillips, C., Haas, C., Pisarek, A., & Lee, H. Y. (December 16, 2021). A collaborative exercise on DNA methylation-based age prediction and body fluid typing. *Forensic Science International: Genetics, 57*, 102656. https://doi.org/10.1016/j.fsigen.2021.102656

Madi, T., Balamurugan, K., Bombardi, R., Duncan, G., & McCord, B. (July 2012). The determination of tissue-specific DNA methylation patterns in forensic biofluids using bisulfite modification and pyrosequencing. *Electrophoresis, 33*(12), 1736−1745. https://doi.org/10.1002/elps.201100711

Mazzone, R., Zwergel, C., Artico, M., Taurone, S., Ralli, M., Greco, A., & Mai, A. (February 26, 2019). The emerging role of epigenetics in human autoimmune disorders. *Clinical Epigenetics, 11*(1), 34. https://doi.org/10.1186/s13148-019-0632-2

Naue, J., Sänger, T., Hoefsloot, H. C. J., Lutz-Bonengel, S., Kloosterman, A. D., & Verschure, P. J. (2018). Proof of concept study of age-dependent DNA methylation markers across different tissues by massive parallel sequencing. *Forensic Science International: Genetics, 36*, 152−159. https://doi.org/10.1016/j.fsigen.2018.07.007

Naue, J., Winkelmann, J., Schmidt, U., & Lutz-Bonengel, S. (2021). Analysis of age-dependent DNA methylation changes in plucked hair samples using massive parallel sequencing. *Rechtsmedizin, 31*, 226−233. https://doi.org/10.1007/s00194-021-00487-4

Park, L. K., Friso, S., & Choi, S. W. (February 2012). Nutritional influences on epigenetics and age-related disease. *Proceedings of the Nutrition Society, 71*(1), 75−83. https://doi.org/10.1017/S0029665111003302

Park, J. L., Kim, J. H., Seo, E., Bae, D. H., Kim, S. Y., Lee, H. C., Woo, K. M., & Kim, Y. S. (July 2016). Identification and evaluation of age-correlated DNA methylation markers for forensic use. *Forensic Science International: Genetics, 23*, 64−70. https://doi.org/10.1016/j.fsigen.2016.03.005

Park, J. L., Kwon, O. H., Kim, J. H., Yoo, H. S., Lee, H. C., Woo, K. M., Kim, S. Y., Lee, S. H., & Kim, Y. S. (November 2014). Identification of body fluid-specific DNA methylation markers for use in forensic science. *Forensic Science International: Genetics, 13*, 147−153. https://doi.org/10.1016/j.fsigen.2014.07.011

Peat, J. R., & Smallwood, S. A. (2018). Low input whole-genome bisulfite sequencing using a post-bisulfite adapter tagging approach. *Methods in Molecular Biology, 1708*, 161−169. https://doi.org/10.1007/978-1-4939-7481-8_9

Pfeifer, M., Bajanowski, T., Helmus, J., & Poetsch, M. (May 2020). Inter-laboratory adaption of age estimation models by DNA methylation analysis-problems and solutions. *International Journal of Legal Medicine, 134*(3), 953−961. https://doi.org/10.1007/s00414-020-02263-7

Rakyan, V. K., Down, T. A., Balding, D. J., & Beck, S. (July 12, 2011). Epigenome-wide association studies for common human diseases. *Nature Reviews Genetics, 12*(8), 529−541. https://doi.org/10.1038/nrg3000

Ramsahoye, B. H., Biniszkiewicz, D., Lyko, F., Clark, V., Bird, A. P., & Jaenisch, R. (May 9, 2000). Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. *Proceedings of the National Academy of Sciences of the U S A, 97*(10), 5237−5242. https://doi.org/10.1073/pnas.97.10.5237

Reed, K., Poulin, M. L., Yan, L., & Parissenti, A. M. (February 1, 2010). Comparison of bisulfite sequencing PCR with pyrosequencing for measuring differences in DNA methylation. *Analytical Biochemistry, 397*(1), 96−106. https://doi.org/10.1016/j.ab.2009.10.021

Romanos, P., & Borjac, J. (2021). Epigenetic discrimination of Lebanese monozygotic twins as a promising forensic approach. *Russian Journal of Genetics, 57*, 90−96. https://doi.org/10.1134/S1022795421010129

Ryan, C. P. (May 2021). "Epigenetic clocks": Theory and applications in human biology. *American Journal of Human Biology, 33*(3), e23488. https://doi.org/10.1002/ajhb.23488

Silva, D. S. B. S., Antunes, J., Balamurugan, K., Duncan, G., Alho, C. S., & McCord, B. (July 2016). Developmental validation studies of epigenetic DNA methylation markers for the detection of blood, semen and saliva samples. *Forensic Science International: Genetics, 23*, 55−63. https://doi.org/10.1016/j.fsigen.2016.01.017

Soares Bispo Santos Silva, D., Antunes, J., Balamurugan, K., Duncan, G., Sampaio Alho, C., & McCord, B. (August 2015). Evaluation of DNA methylation markers and their potential to predict human aging. *Electrophoresis, 36*(15), 1775−1780. https://doi.org/10.1002/elps.201500137

Sukawutthiya, P., Sathirapatya, T., & Vongpaisarnsin, K. (January 2021). A minimal number CpGs of ELOVL2 gene for a chronological age estimation using pyrosequencing. *Forensic Science International, 318*, 110631. https://doi.org/10.1016/j.forsciint.2020.110631

Tammen, S. A., Friso, S., & Choi, S. W. (2013 July–August). Epigenetics: The link between nature and nurture. *Molecular Aspects of Medicine, 34*(4), 753–764. https://doi.org/10.1016/j.mam.2012.07.018

Thong, Z., Chan, X. L. S., Tan, J. Y. Y., Loo, E. S., & Syn, C. K. C. (2017). Evaluation of DNA methylation-based age prediction on blood. *Forensic Science International: Genetics Supplement Series, 6*, e249–e251. https://doi.org/10.1016/J.FSIGSS.2017.09.095

Tytgat, O., Škevin, S., Deforce, D., & Van Nieuwerburgh, F. (January 2022). Nanopore sequencing of a forensic combined STR and SNP multiplex. *Forensic Science International: Genetics, 56*, 102621. https://doi.org/10.1016/j.fsigen.2021.102621

Verogen .com. https://verogen.com/resources/publications/. Accessed on October 29, 2021.

Vidaki, A., Ballard, D., Aliferi, A., Miller, T. H., Barron, L. P., & Syndercombe Court D. (2017). DNA methylation-based forensic age prediction using artificial neural networks and next generation sequencing. *Forensic Science International: Genetics, 28*, 225–236. https://doi.org/10.1016/j.fsigen.2017.02.009

Vidaki, A., Díez López, C., Carnero-Montoro, E., Ralf, A., Ward, K., Spector, T., Bell, J. T., & Kayser, M. (2017). Epigenetic discrimination of identical twins from blood under the forensic scenario. *Forensic Science International: Genetics, 31*, 67–80. https://doi.org/10.1016/j.fsigen.2017.07.014

Vidaki, A., Giangasparo, F., & Syndercombe Court, D. (October 2016). Discovery of potential DNA methylation markers for forensic tissue identification using bisulphite pyrosequencing. *Electrophoresis, 37*(21), 2767–2779. https://doi.org/10.1002/elps.201600261

Watanabe, K., Taniguchi, K., Toyomane, K., & Akutsu, T. (2022). A new approach for forensic analysis of saliva-containing body fluid mixtures based on SNPs and methylation patterns of nearby CpGs. *Forensic Science International: Genetics, 56*, 102624. https://doi.org/10.1016/j.fsigen.2021.102624

Weidner, C. I., Lin, Q., Koch, C. M., Eisele, L., Beier, F., Ziegler, P., Bauerschlag, D. O., Jöckel, K. H., Erbel, R., Mühleisen, T. W., Zenke, M., Brümmendorf, T. H., & Wagner, W. (February 3, 2014). Aging of blood can be tracked by DNA methylation changes at just three CpG sites. *Genome Biology, 15*(2), R24. https://doi.org/10.1186/gb-2014-15-2-r24

Woźniak, A., Heidegger, A., Piniewska-Róg, D., Pośpiech, E., Xavier, C., Pisarek, A., Kartasińska, E., Boroń, M., Freire-Aradas, A., Wojtas, M., de la Puente, M., Niederstätter, H., Płoski, R., Spólnicka, M., Kayser, M., Phillips, C., Parson, W., Branicki, W., & VISAGE Consortium. (2021). Development of the VISAGE enhanced tool and statistical models for epigenetic age estimation in blood, buccal cells and bones. *Aging (Albany NY), 13*(5), 6459–6484. https://doi.org/10.18632/aging.202783

Zbieć-Piekarska, R., Spólnicka, M., Kupiec, T., Makowska, Ż., Spas, A., Parys-Proszek, A., Kucharczyk, K., Płoski, R., & Branicki, W. (2015). Examination of DNA methylation status of the ELOVL2 marker may be useful for human age prediction in forensic science. *Forensic Science International: Genetics, 14*, 161–167. https://doi.org/10.1016/j.fsigen.2014.10.002

Zbieć-Piekarska, R., Spólnicka, M., Kupiec, T., Parys-Proszek, A., Makowska, Ż., Pałeczka, A., Kucharczyk, K., Płoski, R., & Branicki, W. (2015). Development of a forensically useful age prediction method based on DNA methylation analysis. *Forensic Science International: Genetics, 17*, 173–179. https://doi.org/10.1016/j.fsigen.2015.05.001

CHAPTER 18

Forensic applications of NGS-based microRNA analysis

Noora Rashid Al-Snan
Forensic Science Laboratory, Directorate of Forensic Science, General Directorate of Criminal Investigation and Forensic Science, Ministry of Interior, Manama, Bahrain

Introduction

Forensic scientists have long been using nucleic acids (DNA as well as RNA) to produce conclusive evidence in cases of sexual assault, homicide, and all the other cases where any biological evidence is recovered from the victim, suspect, or crime scene (Jawad et al., 2020). With the development of the DNA profiling technique by Sir Alec Jeffreys in 1984, DNA polymorphism gained recognition as a tool for crime investigation. This technique based on the variable number of tandem repeats helped in solving a double murder case (in Leicestershire, UK) by not only identifying the suspect but also saving an innocent (Li, Norman, & Schober, 2015). This demonstrated the huge potential for the use of DNA profiling in forensic and criminal investigations. With the advent of the polymerase chain reaction (PCR) by Kary Mullis, additional molecular techniques were developed for DNA profiling. These techniques include amplified fragment length polymorphism and short tandem repeat (STR) profiling, as well as the more recent single nucleotide polymorphism (SNP), mitochondrial DNA profiling, and Y-STR (Y chromosome-specific STR) profiling.

The DNA profile can help in identifying a person from a biological sample, but it is mostly unable to identify the source, nature, and type of the evidence material. This information can further help to clarify the accusation and support the link between the suspect and the evidence material. It can be reliably done using RNA testing (van den Berge & Sijen, 2017). The RNA testing utilizes the application of different types of RNAs, including mRNA, tRNA, and rRNA, along with miRNA, antisense RNA, siRNA, and snRNA (Jawad et al., 2020). The biological functions of these RNAs may vary from protein biosynthesis to posttranscriptional gene regulation. MicroRNAs (miRNAs) can be defined as short (18–24 nucleotides long) and noncoding RNA molecules that play an important role in gene expression regulation posttranscriptionally. The first miRNA was discovered in the roundworm *Caenorhabditis elegans* in the year 1993. The subsequent research on miRNAs has revealed the crucial regulatory roles of miRNAs in gene expression in plants as well as animals (Lagos-Quintana et al., 2001; Lee & Ambros, 2001). It is estimated that approximately 3% of human genes encode miRNAs.

Next Generation Sequencing (NGS) Technology in DNA Analysis
ISBN 978-0-323-99144-5, https://doi.org/10.1016/B978-0-323-99144-5.00009-3
© 2024 Elsevier Inc.
All rights reserved.

Further, up to 60% of human genes are probable targets of these miRNAs (Git et al., 2010). The data from MiRbase, the bioinformatic database of miRNA sequences, has revealed that 1917 miRNAs are encoded by the human genome (Rocchi et al., 2021). The majority of the miRNAs are found inside the cell, but a few miRNAs have been observed to be circulating in various body fluids, including tears, cerebrospinal fluid, plasma, and saliva. Such extracellular miRNAs are commonly referred to as circulating miRNAs (Sohel, 2016; Weber et al., 2010). The circ miRNAs are potential forensic biomarkers as they have high stability even at high temperatures, extreme pH, and chemical treatments, and as they are carried by the lipoprotein complexes, they are safe from endogenous RNase activity (Layne et al., 2019; Mayes et al., 2019). The advancements in molecular techniques have further helped forensic scientists to develop and validate newer RNA-based methods to solve problems that were unsolvable using conventional methods although various advanced molecular techniques like PCR, microarray, high-resolution melt analysis, etc. have been employed by the researchers (Jawad et al., 2020). Further, next-generation sequencing (NGS) and single-cell sequencing have helped the researchers to develop reliable RNA-based methods for the identification of body fluids, estimation of postmortem interval (PMI), identification of tissue type, determining the age of the wound, distinguishing antemortem and postmortem drowning, gender determination, menstrual blood identification, determining drug abuse, and identifying and discriminating monozygotic twins (Rocchi et al., 2021). The further sections will individually discuss all the aforementioned applications of miRNAs.

miRNA and body fluid identification

Body fluids are one of the most common types of biological evidence found at crime scenes, especially in cases of homicide, suicide, and sexual assault. They may include blood, semen, saliva, urine, and vaginal secretions. The identification of the type of body fluid retrieved from the crime scene is important to determine the nature of the crime, offense reconstruction, relate the suspect to the crime scene, and identify the culprit. Therefore, the very first step in the forensic investigation is to determine the type of body fluid, subsequently followed by DNA profiling to determine the identity of the culprit (Li, Norman, & Schober, 2015). The current presumptive and confirmatory methods used for the identification of the body fluids utilize the different biochemical properties of these fluids to give them color and fluorescence or generate particular compounds. These methods are rapid and show varying levels of sensitivity and specificity. But the fact worthy of mention here is that some of these screening tests are destructive in nature (Harbison & Fleming, 2016; Rocchi et al., 2021; Virkler & Lednev, 2009), which means that a part of the biological evidence will be consumed during these tests. This is a great matter of concern, as most of the time, a small amount of biological evidence is retrieved in criminal cases, and after the preliminary examination, the

same evidence is used for DNA profiling. So, forensic scientists cannot afford to lose part of the evidence. Therefore, it is very important to devise methods that use a small amount of samples and provide confirmatory results.

Alternative methods proposed for body fluid identification (BFID) were DNA methylation profiling and mRNA analysis. But DNA methylation has low specificity, and mRNAs are susceptible to degradation due to cellular, extracellular, and environmental factors (temperature, humidity, etc.) (An et al., 2012; Kader et al., 2020; Lee et al., 2016; Rocchi et al., 2021). This led the researchers to further explore the advanced methods for BFID. Forensic researchers in the last decade have proposed a number of miRNA-based biomarkers for cell-type identification. The initial studies on forensic applications of miRNAs focused on identifying miRNAs that are specific to a body fluid, but most of the miRNAs lacked this specificity (Rocchi et al., 2021). Therefore, comparative miRNA profiling was proposed for better identification of body fluids. The small size of the miRNAs and their feasibility to be coextracted with DNA make the miRNAs an ideal candidate for their use in forensic casework. Two separate studies have reported coanalysis of DNA profiling and BFID. Li, Zhang, et al. (2014) devised a four-miRNA marker-based BFID method that could discriminate between semen, venous blood, and menstrual blood. Along similar lines, Mayes et al. (2018) designed an eight-marker-based system that could differentiate semen, saliva, venous blood, and menstrual blood. Both of these methods yielded complete DNA STR profiles. Individual analysis of miRNA profiling from different body fluids has been performed by various researchers. This analysis was performed using PCR-based methods, microarrays, and NGS. Different authors reported different miRNA biomarkers for BFID, and some of the miRNA show overlapping between different body fluids (Rocchi et al., 2021). The collective list of various miRNAs reported in the literature for BFID has been provided in Table 18.1.

From this table, it can be observed that multiple researchers have reported miR-451a to be a strong biomarker for blood identification, and along with this, miR-144-3p and miR-16 also seem useful biomarkers for venous blood identification. miR-214-3p and miR-412 have higher expression levels in menstrual blood as compared to other body fluids. miR-203a and miR-205-5p have been suggested to be useful in saliva identification. miR-891a, miR-10b, and miR-888 have been confirmed by different researchers as semen-specific makers, and miR-203a has been suggested for vaginal secretion identification.

Discrimination of monozygotic twins

Forensic scientists regularly come across cases that involve the identification of living, deceased, and compromised human remains. Human identification can be scientifically achieved using fingerprint, dental, anthropological, genetic, and radiological examinations (Nikam et al., 2015). With the advent of PCR, DNA fingerprinting became the

Table 18.1 miRNA biomarkers reported in literature for BFID.

Body fluid	Biomarker miRNA	References
Blood (venous)	miR-451a	Courts and Madea (2011), Fujimoto et al. (2019), Hanson et al. (2009), He et al. (2020a, 2020b), Li, Zhang, et al. (2014), Mayes et al. (2018), Omelia et al. (2013), Sirker et al. (2017)
	miR-144-3p	Fujimoto et al. (2019), He, Han, et al. (2020), Sauer et al. (2016), Zubakov et al. (2010)
	miR-16	Fujimoto et al. (2019), Hanson et al. (2009), Wang, Zhang, et al. (2013)
	miR-185	Zubakov et al. (2010)
	miR-126	Courts and Madea (2011)
	miR-150	Courts and Madea (2011)
	miR-486	Wang, Zhang, et al. (2013)
	miR-203a	Sauer et al. (2016)
	miR-943	He, Ji, et al. (2020)
	miR-200b	Seashols-Williams et al. (2016)
	miR-142-3p	Mayes et al. (2018)
Saliva	miR-205-5p	Courts and Madea (2011), Hanson et al. (2009), He, Han, et al. (2020), Mayes et al. (2018), Omelia et al. (2013)
	miR-203a	Courts and Madea (2011), Fujimoto et al. (2019), Hanson et al. (2009), Sauer et al. (2016)
	miR-658	Hanson et al. (2009)
	miR-142-3p	Zubakov et al. (2010)
	miR-200c	Courts and Madea (2011)
	miR-26b	Seashols-Williams et al. (2016)
	miR-223	Fujimoto et al. (2019)
Semen	miR-891a	Belleannée et al. (2012), Fujimoto et al. (2019), He, Han, et al. (2020), Hu et al. (2014), Mayes et al. (2018), Sauer et al. (2016), Seashols-Williams et al. (2016), Wang, Zhang, et al. (2013), Zubakov et al. (2010)
	miR-888	Belleannée et al. (2012), Fujimoto et al. (2019), He, Han, et al. (2020), Hu et al. (2014), Li, Zhang, et al. (2014), Luo et al. (2015), Wang, Zhang, et al. (2013)

Table 18.1 miRNA biomarkers reported in literature for BFID.—cont'd

Body fluid	Biomarker miRNA	References
	miR-10b	Hanson et al. (2009), Mayes et al. (2018), Sirker et al. (2017)
	miR-135b	Hanson et al. (2009)
	miR-135a	Luo et al. (2015), Zubakov et al. (2010)
	miR-374	Sirker et al. (2017)
	miR-10a	Fujimoto et al. (2019)
Menstrual blood	miR-214-3p	He et al. (2020a, 2020b), Li, Zhang, et al. (2014), Wang, Zhang, et al. (2013)
	miR-412	Bexon and Williams (2015), Hanson et al. (2009), Mayes et al. (2018)
	miR-205-5p	Hanson et al. (2009), He, Ji, et al. (2020)
	miR-203a	He, Ji, et al. (2029), Wang et al. (2015)
	miR-144-3p	He, Han, et al. (2020), Sauer et al. (2016)
	miR-451a	Hanson et al. (2009)
	miR-943	Sirker et al. (2017)
	miR-1246	Seashols-Williams et al. (2016)
	miR-200b	Seashols-Williams et al. (2016)
	miR-141-3p	Mayes et al. (2018)
Vaginal secretion	miR-203a	Sauer et al. (2016), Sirker et al. (2017), Wang et al. (2015)
	miR-124a	Hanson et al. (2009)
	miR-372	Hanson et al. (2009)
	miR-200c-3p	Wang et al. (2015)
	miR-205-5p	Wang et al. (2015)
	miR-654-5p	He, Han, et al. (2020)
	miR-888	He, Han, et al. (2020)
	miR-155-5p	Fujimoto et al. (2019)
	miR-1260b	Fujimoto et al. (2019)

most recognized technique for the genetic identification of humans using biological samples. To date, DNA fingerprinting is the most popular, recognized, and scientifically accepted technique for discriminating between two persons from their biological samples in cases involving sexual assault, paternity disputes, and any other case where biological evidence is recovered. This was possible because of the uniqueness of the human genome for every person (Krawczka & Schmidtke, 2020). Despite this, DNA fingerprinting fails to identify and differentiate between monozygotic twins. This is because monozygotic twins have the same genetic makeup. Since monozygotic twins are formed from a single

zygote and are hence genetically identical (Budowle, 2014), their distinction remains a major challenge. Therefore, in cases where one of the monozygotic twins is related to the biological evidence, we cannot exclude the other one. This conundrum of "which one committed the crime?" creates a difficult situation for the legal as well as the forensic communities (Budowle, 2014). To overcome this problem, different methods have been devised for differentiating the monozygotic twins genetically. These methods include copy number variation, SNPs, the mitochondrial genome, and epigenetic-based methods like DNA methylation, RNA expression level, and protein level (TianYue et al., 2020). But, until now, no effective method has been developed that can provide an effective technique for monozygotic twin differentiation.

There are some studies that have reported miRNAs as potential molecular markers for monozygotic twin identification. Perkins et al. (2007), Kim et al. (2010), and Sarachana et al. (2010), in different studies, investigated the expression of miRNAs in monozygotic twins to study the mechanism of diseases and found that monozygotic twins may differ in miRNA expression levels (Kim et al., 2010; Perkins et al., 2007; Sarachana et al., 2010). Chen Fang and colleagues investigated the miRNA expression profile in a pair of monozygotic twins and reported that 96 out of 509 miRNAs were very highly expressed and showed differential expression among monozygotic twins (Fang et al., 2019). Chao Xiao and colleagues studied the miRNA expression profiles of two pairs of monozygotic twins. A total of 74 miRNAs were found to be differentially expressed in the male monozygotic twins, while 220 miRNAs were differentially expressed in the female monozygotic twins (Xiao et al., 2019). Chen Fang et al. analyzed the expression profiles of miRNAs in four pairs of monozygotic twins (Fang et al., 2019). The results showed that out of the 158 miRNAs expressed detectably in each individual, 14% were differentially expressed among monozygotic twins (Fang et al., 2019). Xiao et al. (2019) investigated the miRNA expression in the blood samples of seven pairs of monozygotic twins and found a varying number of differentially expressed miRNAs out of a total of 545 miRNAs among and between monozygotic twins (Xiao et al., 2019).

From these studies, it can be opined that miRNAs are potential genetic markers for differentiating monozygotic twins, but there is a lack of studies on healthy individuals and studies addressing intertissue variation of miRNA profiles. Also, studies have suggested that miRNA expression can be affected by race (Ma, Hong, et al., 2015). Therefore, further studies are needed to establish the feasibility of miRNAs as a genetic marker for monozygotic twin identification.

miRNAs and time since death determination

Accurate estimation of time since death (TSD) is one of the major challenges encountered by investigating agencies during the investigation of deaths. The TSD estimation helps the investigators to determine the time of death, which can further help in

determining whether the case is a suicide or homicide, reconstructing the crime scene, and verifying the alibi of the suspects. The currently used methods for postmortem interval (PMI) estimation are majorly based on physical changes after death. Even the recently developed biochemical methods are unable to provide a precise TSD (Shrestha et al., 2021). Molecular changes in proteins, RNA, and DNA due to postmortem degradation could provide a more precise and reproducible determination of TSD (Li, Ma, et al., 2014; Li et al., 2016). Nucleic acids (DNA and RNA), being protected in the cell nucleus, are considered to be less damaged by external factors and thus may aid in a more precise determination of TSD. The degradation rate of DNA after death has been studied by different researchers, but a detailed study on different human tissues is required for precise PMI estimation (Itani et al., 2011; Johnson & Ferris, 2002; Perez-Martinez et al., 2017; van den Berge et al., 2016; Williams et al., 2015). Scrivano et al. (2019) have provided a detailed review of the studies conducted on the use of RNA in the estimation of the postmortem interval. Most of the studies have been conducted on animal tissues, and only a few studies on human samples have been reported.

Although a lot of research has been conducted on the use of mRNAs for TSD estimation, it is worth noting that the postmortem expression of the mRNAs is highly affected by environmental factors such as temperature, humidity, sunlight, etc. Further, the cause of death as well as the type of tissue concerned also affect postmortem mRNA degradation. The ribonuclease action is higher in the pancreas and liver as compared to the brain, heart, and skeletal muscles. Considering the postmortem instability of the mRNAs, miRNAs can be a potential tool to estimate the precise TSD. Forensic researchers have shifted their focus toward using miRNAs as a potential biomarker for TSD estimation. Lesser susceptibility to degradation has led to great attention toward miRNAs. As the miRNAs are smaller in size as compared to mRNAs, they are more likely to be found intact postcease of life. Further, miRNAs are protected by lipid or protein matrices, which add another layer of protection from endonucleases and ribonucleases (Rocchi et al., 2021). Several researchers have reported the stability of miRNAs even after treatment with varying environmental conditions. Layne et al. (2019) reported four miRNAs that were resistant to UV light exposure in blood, semen, saliva, and urine samples. Similarly, two miRNAs were found to be stable in blood and semen even after 6 months of differential environmental conditions of temperature, humidity, and sunlight (Layne et al., 2019; Mayes et al., 2019).

Li et al. (2010) were the first researchers to report the potential use of miRNAs for TSD estimation. Although several studies have reported no correlation between miRNA expression and TSD in the early postmortem interval (0−24 h), a number of researchers have reported a positive correlation of a number of miRNAs with TSD in the late (24−72 h) and advanced stages (>72 h) of the postmortem interval (Rocchi et al., 2021). miR-1-2 miRNA markers were identified in rat cardiac tissue for their potential use in TSD estimation from 96 to 120 h after death. Further, two miRNAs, namely

miR-206 and miR-133, from mouse liver tissue were suggested to be useful from 24 to 48 h after death. The study on human bone samples revealed that let-7e and miR-16c can be used as potential biomarkers for TSD estimation in advanced stages of decomposition. Few researchers have investigated the expression levels of miRNAs regulating the circadian clock during the day and night to determine the time of death. Expression levels of four miRNA markers in human vitreous humor samples were found to be linked with daytime and nighttime. miR-142-5p and miR-541 had lower expression in the daytime and higher expression at nighttime. However, miR-106b and miR-96 had higher expression during the daytime and lower expression at nighttime. miR-142-5p and miR-219 from human blood had higher expression during the daytime and lower expression at nighttime (Rocchi et al., 2021).

As the expressional analysis of the RNAs is done using PCR, it is important to set suitable control biomarkers. As discussed, mRNAs are highly unstable, and this renders it difficult to use them as control markers while studying the expression levels in postmortem tissues. Therefore, researchers explored the possibility of using miRNAs as reference or control biomarkers in postmortem studies. Various studies were conducted on different tissues from humans, rats, and mice, and a number of miRNAs were reported to be ideal for reference biomarkers (Rocchi et al., 2021). As miRNA expression varies among tissue types, the reference miRNAs should be selected as per the tissue under analysis. A comprehensive list of the miRNAs reported in the literature to be linked with TSD has been provided in Table 18.2.

It can be concluded that miRNA expression is associated with TSD in late PMI (24−144 h after death), and there's relatively no correlation between miRNA expression and TSD in early PMI (0−24 h after death). It should be further noted that there is a need to increase the number of studies on human tissue samples with higher sample sizes.

Wound age estimation using miRNA

Medico-legal examination of wounds is necessary to determine the severity of the wounds. At the same time, it is also important to determine the age of the wound or the time since the wound was inflicted, as well as whether the wounds were inflicted antemortem or postmortem. It helps the investigators to verify the statements of the alibi and reconstruct the crime's events. When an injury is inflicted, the body immediately starts the healing process. The healing process has been divided into four phases, which overlap each other. The four phases are the hemostasis, inflammatory, proliferative, and remodeling phases. Each phase has unique morphological and biochemical properties. The detection of the healing phase-specific characteristics has helped the researchers to determine the age of the wound. Various immunohistochemical markers, cytokines, chemokines, adhesion molecules, different enzymes, and growth factors have been analyzed for determining the age of the wound (Casse et al., 2016).

Table 18.2 miRNA markers reported in the literature for TSD estimation.

Application	Source of tissue	Tissue type	miRNA	Postmortem stability period	References
TSD estimation	Human	Bones	Let-7e and miR-16	1–6 months	Na (2020)
	Rat	Heart	miR-1-2	Up to 120 h	Li et al. (2010)
	Mouse	Liver	miR-133a and miR-206	24–48 h	Wang, Mao, et al. (2013)
Determining circadian rhythm	Human	Vitreous humor	miR-142-5p and miR-541	Low daytime high nighttime	Odriozola et al. (2013)
			miR-106b and miR-96	High daytime low nighttime	Corradini et al. (2015)
		Blood	miR-142-5p and miR-219	High daytime low nighttime	Corradini et al. (2015)
For internal control/ Reference marker in PCR	Human	Lung	miR-195 and miR-200c	Up to 144 h	Lv et al. (2016)
		Muscle	miR-1 and miR-206	Up to 144 h	Lv et al. (2016)
		Brain	miR-9, miR-125b	Up to 22 h	Lü, Ma, et al. (2016)
		Heart, liver, brain	miR-1 and miR-133a	Up to 180 h	Lv et al. (2017)
	Rat	Skin	miR-203	Up to 120 h	Pan et al. (2014)
		Spleen	miR-125b and miR-143	Up to 36 h (25°C) Up to 144 h (4°C)	Lv et al. (2014)
		Lung	miR-195 and miR-200c	Up to 144 h	Lv et al. (2016)
		Muscle	miR-1 and miR-206	Up to 144 h	Lv et al. (2016)
		Brain	miR-9 and miR-125b	Up to 144 h	Ma, Pan, et al. (2015)
				Up to 24 h	Lü, Li, et al. (2016)
		Heart, liver	miR-1 and miR-133a	Up to 180 h	Lv et al. (2017)
	Mouse	Heart	miR-122 and miR-133a	Up to 180 h	Tu et al. (2018, 2019)
		Liver	miR-122	Up to 180 h	Tu et al. (2018)
		Skeletal muscle	miR-133a	Up to 180 h	Tu et al. (2018)

Specific miRNAs have been reported to regulate the genes involved in the wound-healing process. This makes miRNAs a potential biomarker to estimate the age of wounds (Rocchi et al., 2021). Different miRNAs have been reported to be unique to the wound-healing phase. Table 18.3 provides a list of the miRNAs identified to be involved in various phases of wound healing.

There is a paucity of literature on differences in antemortem and postmortem expression profiles of miRNAs. Only a few studies have been published in this area of research. miRNA-21 and miRNA-205 have been reported to be significantly up-regulated in rat skin samples at 24 h after death as compared to 0 and 48 h after death. However, no correlation could be established between the TSD and miRNA expression levels (Ibrahim et al., 2019). Also, increased levels of two miRNAs involved in the inflammatory phase, namely miR125a-5p and miR125b-5p have been reported in the skin ligature marks of individuals who died by hanging. Further, increased levels of three miRNAs with an antiinflammatory role, miR92a-3p, miR128-3p, and miR130a-3p, were observed (Neri et al., 2019).

It can be concluded that although a few studies have shown the potential candidature of miRNAs as biomarkers for wound age estimation but more research is required to establish miRNAs for practical applications for this purpose.

Organ tissue identification using miRNA

Internal organ tissues may be retrieved in cases of violent crimes such as stabbing and shooting. It may be found on the crime scene, a weapon used, or the body of the perpetrator. The identification of the type of organ tissue can help in determining the level or depth of weapon penetration, the role of the evidence material on which the tissue was found, and in certain cases, if human remains have been dismembered, disposed of, or

Table 18.3 miRNAs identified to be involved in various phases of wound healing.

miRNA	Phase of wound healing	References
miR-21, miR-146a/b, miR-142, and miR-155	Inflammatory phase	Li et al. (2018), Mori et al. (2018), Roy and Sen (2011), Santurro et al. (2017)
miR-21 and miR-132	Inflammatory and proliferative phase	Li, Wang, et al. (2015), Madhyastha et al. (2012), Yang et al. (2011)
miR-31 and miR-99	Proliferative phase	Jin et al. (2013), Li, Li, et al. (2015)
miR-29a,b,c and miR-192	Remodeling phase	Cheng et al. (2010), Ciechomska et al. (2014)

transferred to another location. Currently, no standardized molecular methods exist for the identification of organ tissues. Forensic scientists have to rely on immunohistochemical methods for tissue identification. Such methods require a good amount of tissue samples in pristine conditions. But the tissues found in criminal cases are usually degraded and only in trace amounts. Some researchers reported promising methods to identify tissue organs based on mRNA profiling and DNA methylation levels, but the postmortem instability of these molecules renders it difficult to put these methods to practical use.

miRNAs have been shown to be potential biomarkers for BFID, and therefore their potential should be explored for tissue identification as well. In one such research, Sauer et al. (2017) found that tissues from the brain, heart muscle, lung, liver, skin, and skeletal muscle could be identified based on the expression levels of 15 miRNAs. They also validated this study on samples from mock shootings and stabbings. This study identified individual miRNAs that were differentially expressed in the different tissues and thus could be used for tissue identification. It was noted that a combination of multiplexed miRNAs should be used for better inferences. This study provides a strong foundation, and further studies should be conducted with an increased sample size and more tissue types (Sauer et al., 2017).

miRNA and drowning

Drowning cases are commonly encountered during forensic investigations. The two major questions that need to be answered in drowning cases are whether the death really occurred due to drowning and whether the drowning was postmortem or antemortem. Currently, the methods employed for such investigations include the analysis of diatoms in the deceased tissues, electrolyte differences in the blood, pleural liquid, and vitreous humor of the deceased, along with the evaluation of aquaporins and some other biomarkers. Microenvironmental changes such as oxygen saturation and blood electrolyte concentration during drowning can modify the ion-transport system of various cells. Certain miRNAs have been reported to regulate the genes that control ion channels (Sepramaniam et al., 2010). The microenvironmental changes that happen at the time of death can provide lead to miRNA expression patterns and thus aid in determining the type of drowning.

Yu et al. (2015) conducted a study to analyze the miRNA expression pattern in the brain tissue of drowning animal models. The findings of this study revealed that there were 158 differentially expressed miRNAs, but only eight miRNAs could be established as significantly upregulated in freshwater drowning and downregulated in saltwater drowning as compared to the control samples. Further analysis showed that one of these eight miRNAs, "miRNA-706," is involved in the regulation of the HCN1 cation channel. Thus, miRNA-706 could be a biomarker to differentiate between freshwater and saltwater drowning (Yu et al., 2015). Although this study has shown the fundamental

behavior of a few miRNAs during drowning, this is the only reported study in this area. Therefore, more research is required to find candidate miRNA biomarkers applicable to drowning cases.

miRNA and sepsis

Sepsis refers to a life-threatening condition that is caused by a poorly regulated inflammatory response of the immune system. It has a high mortality rate (Dellinger et al., 2013). The pathogenesis of sepsis is still not well understood. The postmortem diagnosis of sepsis is difficult as there is no specific histological evidence found (Tsokos, 2007). miRNAs have a key role in the regulation of genes responsible for inflammatory responses, and therefore researchers have suggested the use of miRNAs as sepsis biomarkers. Reithmair et al. (2017) have proposed a set of 180 miRNAs for the precise assessment of sepsis-related deaths. They analyzed blood samples of 22 human patients with sepsis along with 23 control samples. The miRNA expression profiles were investigated in exosomes, serum, and blood cells. The findings revealed that there were 103 miRNAs with higher levels of expression and 77 miRNAs with lower levels of expression in sepsis patients as compared to the control samples. Further, the authors identified miR-199b-5p as a potential early biomarker that can be useful in distinguishing patients with sepsis (Reithmair et al., 2017).

It has also been observed that circulating miRNAs could signify the presence of particular pathogens. Identification of particular pathogens can help in the investigation of sepsis negligence cases. Seven miRNAs have been reported to be correlated with infection due to *Staphylococcus aureus*, and eight miRNAs have been reported to be correlated with infection due to *Escherichia coli* (Wu et al., 2013). In addition to these pathogens, approximately 10 other pathogen infections have also been investigated for differential miRNA profiles (Rocchi et al., 2021).

A clearer view of the application of miRNA in sepsis assessment can only be obtained after thorough research with a larger sample size has been conducted in this area.

Other potential future application

No doubt, miRNAs are the newfound love of biomedical as well as forensic researchers, but most of the forensic research in this area is at the preliminary stage only. A number of potential forensic applications of miRNAs have been suggested, and research is going on most of them. Other than the aforementioned areas of forensics, miRNA can be a potential tool for the investigation of fire injuries. Some interesting findings have been reported that suggest that miRNAs can be used as a promising biomarker to distinguish antemortem and postmortem burns (Lyu et al., 2018). As the administration of any drug can alter the expression profiles of miRNAs, miRNAs can be used for antidoping

purposes. Various authors have investigated the correlation of miRNAs with drug and substance abuse and found promising results (Rocchi et al., 2021). But extensive studies with larger cohorts are required to find specific miRNA signatures related to drug abuse. Further, the age-related changes in miRNA expression profiles can be used to determine the age of the person whose biological sample is under analysis. Similarly, the time since the deposition of a biological stain can also be determined based on miRNA profiling (Glynn, 2020). Meticulous future research is required to investigate these promising forensic applications of miRNAs.

Potential limitations

The routine application of miRNA-based techniques in forensic labs is still debatable, as this is an emerging field that needs a lot of standardization. The PCR-based techniques involve the complexity of appropriate primer design, and the cross-reactivity of the markers puts a limit on the routine use of this promising approach. The lack of agreement in the published literature further makes it difficult to bring this approach into routine use. It can be due to differential methodological as well as analytical approaches being used by various researchers across the globe. Further, the commercial unavailability of bio-analytical kits specifically designed for forensic purposes also poses a difficulty for the researchers (Glynn, 2020). Another important issue in miRNA analysis is the statistical analysis of the data. It is yet to be standardized as to which statistical approach is best suited to determine whether two datasets are significantly different or not (Rocchi et al., 2021).

Concluding remarks

The forensic techniques that once seemed advanced are becoming antiquated in front of the newer challenges. Modern problems require modern solutions, and miRNA is one such solution to a lot of investigation problems encountered by forensic scientists. Although a major focus of miRNA research across the globe has been limited to its biomedical and clinical applications, recent literature has suggested a plethora of applications in the forensic field. Being recently recognized and still in its emerging state, the research on the forensic applications of miRNA is limited, but it's growing continuously. Detailed research is required on each application of miRNA to devise robust techniques that can produce precise results acceptable to forensic scientists as well as the court.

Acknowledgments

We would like to thank Osama Hamoudi for his support in writing this chapter.

References

An, J. H., Shin, K.-J., Yang, W. I., & Lee, H. Y. (2012). Body fluid identification in forensics. *BMB Reports, 45*(10), 545–553. https://doi.org/10.5483/BMBRep.2012.45.10.206

Belleannée, C., Calvo, E., Thimon, V., Cyr, D. G., Légaré, C., Garneau, L., & Sullivan, R. (2012). Role of microRNAs in controlling gene expression in different segments of the human epididymis. *PLoS One,* 7, e34996. https://doi.org/10.1371/journal.pone.0034996

Bexon, K., & Williams, G. (2015). Characterising the fluctuation of microRNA expression throughout a full menstrual cycle. *Forensic Science International: Genetics Supplement Series, 5*, e264–e266. https://doi.org/10.1016/j.fsigss.2015.09.105

Budowle, B. (2014). Editors' pick: Molecular genetic investigative leads to differentiate monozygotic twins. *Investigative Genetics, 5*, 1–2.

Casse, J.-M., Martrille, L., Vignaud, J.-M., & Gauchotte, G. (2016). Skin wounds vitality markers in forensic pathology: An updated review. *Medicine, Science & the Law, 56*, 128–137. https://doi.org/10.1177/0025802415590175

Cheng, J., Yu, H., Deng, S., & Shen, G. (2010). MicroRNA profiling in mid-and late-gestational fetal skin: Implication for scarless wound healing. *Tohoku Journal of Experimental Medicine, 221*, 203–209. https://doi.org/10.1620/tjem.221.203

Ciechomska, M., O'Reilly, S., Suwara, M., Bogunia-Kubik, K., & van Laar, J. M. (2014). MiR-29a reduces TIMP-1 production by dermal fibroblasts via targeting TGF-β activated kinase 1 binding protein 1, implications for systemic sclerosis. *PLoS One, 9*, e115596. https://doi.org/10.1371/journal.pone.0115596

Corradini, B., Alù, M., Radheshi, E., Gabbolini, V., Ferrari, F., Santunione, A. L., & Silingardi, E. (2015). Estimation of the time of death through the analysis of clock miRNAs expression in blood and vitreous humor. *Forensic Science International: Genetics Supplement Series, 5*, e204–e206. https://doi.org/10.1016/j.fsigss.2015.09.082

Courts, C., & Madea, B. (2011). Specific micro-RNA signatures for the detection of saliva and blood in forensic body-fluid identification. *Journal of Forensic Sciences, 56*, 1464–1470. https://doi.org/10.1111/j.1556-4029.2011.01894.x

Dellinger, R. P., Levy, M. M., Rhodes, A., Annane, D., Gerlach, H., Opal, S. M., Sevransky, J. E., Sprung, C., Douglas, I., & Jaeschke, R. (2013). Surviving sepsis campaign guidelines committee including the pediatric subgroup surviving sepsis campaign: International guidelines for management of severe sepsis and septic shock: 2012. *Critical Care Medicine, 41*, 580–637. https://doi.org/10.1007/s00134-012-2769-8

Fang, C., Zhao, J., Liu, X., Zhang, J., Cao, Y., Yang, Y., Yu, C., Zhang, X., Qian, J., & Liu, W. (2019). MicroRNA profile analysis for discrimination of monozygotic twins using massively parallel sequencing and real-time PCR. *Forensic Science International: Genetics, 38*, 23–31. https://doi.org/10.1016/j.fsigen.2018.09.011

Fujimoto, S., Manabe, S., Morimoto, C., Ozeki, M., Hamano, Y., Hirai, E., Kotani, H., & Tamaki, K. (2019). Distinct spectrum of microRNA expression in forensically relevant body fluids and probabilistic discriminant approach. *Scientific Reports, 9*, 1–10. https://doi.org/10.1038/s41598-019-50796-8

Git, A., Dvinge, H., Salmon-Divon, M., Osborne, M., Kutter, C., Hadfield, J., Bertone, P., & Caldas, C. (2010). Systematic comparison of microarray profiling, real-time PCR, and next-generation sequencing technologies for measuring differential microRNA expression. *RNA, 16*, 991–1006. https://doi.org/10.1261/rna.1947110

Glynn, C. L. (2020). Potential applications of microRNA profiling to forensic investigations. *RNA, 26*, 1–9. https://doi.org/10.1261/rna.072173.119

Hanson, E. K., Lubenow, H., & Ballantyne, J. (2009). Identification of forensically relevant body fluids using a panel of differentially expressed microRNAs. *Analytical Biochemistry, 387*, 303–314. https://doi.org/10.1016/j.ab.2009.01.037

Harbison, S., & Fleming, R. (2016). Forensic body fluid identification: State of the art. *Research and Reports in Forensic Medical Science, 6*, 11–23. https://doi.org/10.2147/RRFMS.S57994

He, H., Han, N., Ji, C., Zhao, Y., Hu, S., Kong, Q., Ye, J., Ji, A., & Sun, Q. (2020). Identification of five types of forensic body fluids based on stepwise discriminant analysis. *Forensic Science International: Genetics, 48*, 102337. https://doi.org/10.1016/j.fsigen.2020.102337

He, H., Ji, A., Zhao, Y., Han, N., Hu, S., Kong, Q., Jiang, L., Ye, J., Liu, Y., & Sun, Q. (2020). A stepwise strategy to distinguish menstrual blood from peripheral blood by Fisher's discriminant function. *International Journal of Legal Medicine, 134*, 845–851. https://doi.org/10.1007/s00414-019-02196-w

Hu, L., Wu, C., Guo, C., Li, H., & Xiong, C. (2014). Identification of microRNAs predominately derived from testis and epididymis in human seminal plasma. *Clinical Biochemistry, 47*, 967–972. https://doi.org/10.1016/j.clinbiochem.2013.11.009

Ibrahim, S. F., Ali, M. M., Basyouni, H., Rashed, L. A., Amer, E. A., & Abd El-Kareem, D. (2019). Histological and miRNAs postmortem changes in incisional wound. *Egyptian Journal of Food Science, 9*, 1–6. https://doi.org/10.1186/s41935-019-0141-7

Itani, M., Yamamoto, Y., Doi, Y., & Miyaishi, S. (2011). Quantitative analysis of DNA degradation in the dead body. *Acta Medicinae Okayama, 65*, 299–306. https://doi.org/10.18926/AMO/47011

Jawad, M., Noreen, S., Asif, M. F. A. M., Asif, H. M., Ali, S., & Rakha, A. (2020). Application of ribonucleic acid (RNA) in forensic sciences. *Biomedica, 36*. https://doi.org/10.51441/biomedica/5-101

Jin, Y., Tymen, S. D., Chen, D., Fang, Z. J., Zhao, Y., Dragas, D., Dai, Y., Marucha, P. T., & Zhou, X. (2013). MicroRNA-99 family targets AKT/mTOR signaling pathway in dermal wound healing. *PLoS One, 8*, e64434. https://doi.org/10.1371/journal.pone.0064434

Johnson, L. A., & Ferris, J. A. (2002). Analysis of postmortem DNA degradation by single-cell gel electrophoresis. *Forensic Science International, 126*, 43–47. https://doi.org/10.1016/s0379-0738(02)00027-0

Kader, F., Ghai, M., & Olaniran, A. O. (2020). Characterization of DNA methylation-based markers for human body fluid identification in forensics: A critical review. *International Journal of Legal Medicine, 134*, 1–20. https://doi.org/10.1007/s00414-019-02181-3

Kim, A. H., Reimers, M., Maher, B., Williamson, V., McMichael, O., McClay, J. L., van den Oord, E. J., Riley, B. P., Kendler, K. S., & Vladimirov, V. I. (2010). MicroRNA expression profiling in the prefrontal cortex of individuals affected with schizophrenia and bipolar disorders. *Schizophrenia Research, 124*, 183–191. https://doi.org/10.1016/j.schres.2010.07.002

Krawczka, M., & Schmidtke, J. (2020). *DNA fingerprinting*. Taylor & Francis.

Lagos-Quintana, M., Rauhut, R., Lendeckel, W., & Tuschl, T. (2001). Identification of novel genes coding for small expressed RNAs. *Science, 294*, 853–858. https://doi.org/10.3410/f.1002346.20104

Layne, T. R., Green, R. A., Lewis, C. A., Nogales, F., Dawson Cruz, T. C., Zehner, Z. E., & Seashols-Williams, S. J. (2019). Micro RNA detection in blood, urine, semen, and saliva stains after compromising treatments. *Journal of Forensic Sciences, 64*, 1831–1837. https://doi.org/10.1111/1556-4029.14113

Lee, R. C., & Ambros, V. (2001). An extensive class of small RNAs in *Caenorhabditis elegans*. *Science, 294*, 862–864. https://doi.org/10.1126/science.1065329

Lee, H. Y., Jung, S.-E., Lee, E. H., Yang, W. I., & Shin, K.-J. (2016). DNA methylation profiling for a confirmatory test for blood, saliva, semen, vaginal fluid and menstrual blood. *Forensic Science International: Genetics, 24*, 75–82. https://doi.org/10.1016/j.fsigen.2016.06.007

Li, X., Guo, L., Liu, Y., Su, Y., Xie, Y., Du, J., Wang, S., Wang, H., & Liu, Y. (2018). MicroRNA-21 promotes wound healing via the Smad7-Smad2/3-Elastin pathway. *Experimental Cell Research, 362*, 245–251. https://doi.org/10.1016/j.yexcr.2017.11.019

Li, D., Li, X., Wang, A., Meisgen, F., Pivarcsi, A., Sonkoly, E., Ståhle, M., & Landén, N. X. (2015). MicroRNA-31 promotes skin wound healing by enhancing keratinocyte proliferation and migration. *Journal of Investigative Dermatology, 135*, 1676–1685. https://doi.org/10.1038/jid.2015.48

Li, W. C., Ma, K. J., Lv, Y. H., Zhang, P., Pan, H., Zhang, H., Wang, H. J., Ma, D., & Chen, L. (2014). Postmortem interval determination using 18S-rRNA and microRNA. *Science & Justice, 54*, 307–310. https://doi.org/10.1016/j.scijus.2014.03.001

Li, W., Ma, K., Zhang, P., Wang, H., Shen, Y., Zhou, Y., Zhao, Z., Ma, D., & Chen, L. (2010). Estimation of postmortem interval using microRNA and 18S rRNA degradation in rat cardiac muscle. *Fa Yi Xue Za Zhi, 26*, 413–417.

Li, R., Norman, S., & Schober, J. (2015). *Forensic biology*. CRC Press.

Li, D., Wang, A., Liu, X., Meisgen, F., Grünler, J., Botusan, I. R., Narayanan, S., Erikci, E., Li, X., & Blomqvist, L. (2015). MicroRNA-132 enhances transition from inflammation to proliferation during wound healing. *The Journal of Clinical Investigation, 125*, 3008–3026. https://doi.org/10.1172/jci79052

Li, C., Wang, Q., Zhang, Y., Lin, H., Zhang, J., Huang, P., & Wang, Z. (2016). Research progress in the estimation of the postmortem interval by Chinese forensic scholars. *Forensic Sciences Research, 1*, 3−13. https://doi.org/10.1080/20961790.2016.1229377

Li, Y., Zhang, J., Wei, W., Wang, Z., Prinz, M., & Hou, Y. (2014). A strategy for co-analysis of microRNAs and DNA. *Forensic Science International: Genetics, 12*, 24−29. https://doi.org/10.1016/j.fsigen.2014.04.011

Lü, Y., Li, Z., Tuo, Y., Liu, L., Li, K., Bian, J., Ma, J., & Chen, L. (2016). Correlation between RNA degradation patterns of rat's brain and early PMI at different temperatures. *Fa Yi Xue Za Zhi, 32*, 165−170.

Lü, Y., Ma, K., Li, Z., Gu, J., Bao, J., Yang, Z., Gao, J., Zeng, Y., Tao, L., & Chen, L. (2016). Correlation between RNA expression level and early PMI in human brain tissue. *Fa Yi Xue Za Zhi, 32*, 245−249.

Luo, X., Li, Z., Peng, D., Wang, L., Zhang, L., & Liang, W. (2015). MicroRNA markers for forensic body fluid identification obtained from miRCURYTM LNA array. *Forensic Science International: Genetics Supplement Series, 5*, e630−e632. https://doi.org/10.1016/j.fsigss.2015.10.006

Lv, Y.-H., Ma, J.-L., Pan, H., Zeng, Y., Tao, L., Zhang, H., Li, W.-C., Ma, K.-J., & Chen, L. (2017). Estimation of the human postmortem interval using an established rat mathematical model and multi-RNA markers. *Forensic Science, Medicine and Pathology, 13*, 20−27. https://doi.org/10.1007/s12024-016-9827-4

Lv, Y.-H., Ma, J.-L., Pan, H., Zhang, H., Li, W.-C., Xue, A.-M., Wang, H.-J., Ma, K.-J., & Chen, L. (2016). RNA degradation as described by a mathematical model for postmortem interval determination. *Journal of Forensic and Legal Medicine, 44*, 43−52. https://doi.org/10.1016/j.jflm.2016.08.015

Lv, Y. H., Ma, K. J., Zhang, H., He, M., Zhang, P., Shen, Y. W., Jiang, N., Ma, D., & Chen, L. (2014). A time course study demonstrating mRNA, microRNA, 18S rRNA, and U6 snRNA changes to estimate PMI in deceased rat's spleen. *Journal of Forensic Sciences, 59*, 1286−1294. https://doi.org/10.1111/1556-4029.12447

Lyu, H.-P., Cheng, M., Liu, J.-C., Ye, M.-Y., Xu, D., He, J.-T., Xie, X.-L., & Wang, Q. (2018). Differentially expressed microRNAs as potential markers for vital reaction of burned skin. *Journal of Forensic Science and Medicine, 4*, 135. https://doi.org/10.4103/jfsm.jfsm_1_18

Madhyastha, R., Madhyastha, H., Nakajima, Y., Omura, S., & Maruyama, M. (2012). MicroRNA signature in diabetic wound healing: Promotive role of miR-21 in fibroblast migration. *International Wound Journal, 9*, 355−361. https://doi.org/10.1111/j.1742-481x.2011.00890.x

Ma, L., Hong, Y., Lu, C., Chen, Y., & Ma, C. (2015). The occurrence of cervical cancer in Uygur women in Xinjiang Uygur autonomous region is correlated to microRNA-146a and ethnic factor. *International Journal of Clinical and Experimental Pathology, 8*, 9368. 26464690.

Ma, J., Pan, H., Zeng, Y., Lv, Y., Zhang, H., Xue, A., Jiang, J., Ma, K., & Chen, L. (2015). Exploration of the R code-based mathematical model for PMI estimation using profiling of RNA degradation in rat brain tissue at different temperatures. *Forensic Science, Medicine and Pathology, 11*, 530−537. https://doi.org/10.1007/s12024-015-9703-7

Mayes, C., Houston, R., Seashols-Williams, S., LaRue, B., & Hughes-Stamm, S. (2019). The stability and persistence of blood and semen mRNA and miRNA targets for body fluid identification in environmentally challenged and laundered samples. *Legal Medicine, 38*, 45−50. https://doi.org/10.1016/j.legalmed.2019.03.007

Mayes, C., Seashols-Williams, S., & Hughes-Stamm, S. (2018). A capillary electrophoresis method for identifying forensically relevant body fluids using miRNAs. *Legal Medicine, 30*, 1−4. https://doi.org/10.1016/j.legalmed.2017.10.013

Mori, R., Tanaka, K., & Shimokawa, I. (2018). Identification and functional analysis of inflammation-related mi RNAs in skin wound repair. *Development Growth and Differentiation, 60*, 306−315. https://doi.org/10.1111/dgd.12542

Na, J. Y. (2020). Estimation of the post-mortem interval using microRNA in the bones. *Journal of Forensic and Legal Medicine, 75*, 102049. https://doi.org/10.1016/j.jflm.2020.102049

Neri, M., Fabbri, M., D'Errico, S., Di Paolo, M., Frati, P., Gaudio, R. M., La Russa, R., Maiese, A., Marti, M., & Pinchi, E. (2019). Regulation of miRNAs as new tool for cutaneous vitality lesions demonstration in ligature marks in deaths by hanging. *Scientific Reports, 9*, 1−10. https://doi.org/10.1038/s41598-019-56682-7

Nikam, S. S., Gadgil, R. M., Bhoosreddy, A. R., Shah, K. R., & Shirsekar, V. U. (2015). Personal identification in forensic science using uniqueness of radiographic image of frontal sinus. *The Journal of Forensic Odonto-Stomatology, 33*, 1.

Odriozola, A., Riancho, J. A., de la Vega, R., Agudo, G., García-Blanco, A., de Cos, E., Fernández, F., Sañudo, C., & Zarrabeitia, M. T. (2013). miRNA analysis in vitreous humor to determine the time of death: A proof-of-concept pilot study. *International Journal of Legal Medicine, 127*, 573–578. https://doi.org/10.1007/s00414-012-0811-6

Omelia, E. J., Uchimoto, M. L., & Williams, G. (2013). Quantitative PCR analysis of blood-and saliva-specific microRNA markers following solid-phase DNA extraction. *Analytical Biochemistry, 435*, 120–122. https://doi.org/10.1016/j.ab.2012.12.024

Pan, H., Zhang, H., Lü, Y., Ma, J., Ma, K., & Chen, L. (2014). Correlation between five RNA markers of rat's skin and PMI at different temperatures. *Fa Yi Xue Za Zhi, 30*, 245–249. 25434083.

Perez-Martinez, C., Perez-Carceles, M. D., Legaz, I., Prieto-Bonete, G., & Luna, A. (2017). Quantification of nitrogenous bases, DNA and Collagen type I for the estimation of the postmortem interval in bone remains. *Forensic Science International, 281*, 106–112. https://doi.org/10.1016/j.forsciint.2017.10.039

Perkins, D. O., Jeffries, C. D., Jarskog, L. F., Thomson, J. M., Woods, K., Newman, M. A., Parker, J. S., Jin, J., & Hammond, S. M. (2007). MicroRNA expression in the prefrontal cortex of individuals with schizophrenia and schizoaffective disorder. *Genome Biology, 8*, 1–11. https://doi.org/10.1186/gb-2007-8-2-r27

Reithmair, M., Buschmann, D., Märte, M., Kirchner, B., Hagl, D., Kaufmann, I., Pfob, M., Chouker, A., Steinlein, O. K., & Pfaffl, M. W. (2017). Cellular and extracellular mi RNAs are blood-compartment-specific diagnostic targets in sepsis. *Journal of Cellular and Molecular Medicine, 21*, 2403–2411. https://doi.org/10.1111/jcmm.13162

Rocchi, A., Chiti, E., Maiese, A., Turillazzi, E., & Spinetti, I. (2021). MicroRNAs: An update of applications in forensic science. *Diagnostics, 11*, 32. https://doi.org/10.3390/diagnostics11010032

Roy, S., & Sen, C. K. (2011). MiRNA in innate immune responses: Novel players in wound inflammation. *Physiological Genomics, 43*, 557–565. https://doi.org/10.1152/physiolgenomics.00160.2010

Santurro, A., Maria Vullo, A., Borro, M., Gentile, G., La Russa, R., Simmaco, M., Frati, P., & Fineschi, V. (2017). Personalized medicine applied to forensic sciences: New advances and perspectives for a tailored forensic approach. *Current Pharmaceutical Biotechnology, 18*, 263–273. https://doi.org/10.2174/1389201018666170207141525

Sarachana, T., Zhou, R., Chen, G., Manji, H. K., & Hu, V. W. (2010). Investigation of post-transcriptional gene regulatory networks associated with autism spectrum disorders by microRNA expression profiling of lymphoblastoid cell lines. *Genome Medicine, 2*, 1–18. https://doi.org/10.1186/gm144

Sauer, E., Extra, A., Cachée, P., & Courts, C. (2017). Identification of organ tissue types and skin from forensic samples by microRNA expression analysis. *Forensic Science International: Genetics, 28*, 99–110. https://doi.org/10.1016/j.fsigen.2017.02.002

Sauer, E., Reinke, A.-K., & Courts, C. (2016). Differentiation of five body fluids from forensic samples by expression analysis of four microRNAs using quantitative PCR. *Forensic Science International: Genetics, 22*, 89–99. https://doi.org/10.1016/j.fsigen.2016.01.018

Seashols-Williams, S., Lewis, C., Calloway, C., Peace, N., Harrison, A., Hayes-Nash, C., Fleming, S., Wu, Q., & Zehner, Z. E. (2016). High-throughput miRNA sequencing and identification of biomarkers for forensically relevant biological fluids. *Electrophoresis, 37*, 2780–2788. https://doi.org/10.1002/elps.201600258

Sepramaniam, S., Armugam, A., Lim, K. Y., Karolina, D. S., Swaminathan, P., Tan, J. R., & Jeyaseelan, K. (2010). MicroRNA 320a functions as a novel endogenous modulator of aquaporins 1 and 4 as well as a potential therapeutic target in cerebral ischemia*[S]. *Journal of Biological Chemistry, 285*, 29223–29230. https://doi.org/10.1074/jbc.m110.144576

Shrestha, R., Kanchan, T., & Krishan, K. (2021). *Methods of estimation of time since death*. StatPearls.

Sirker, M., Fimmers, R., Schneider, P., & Gomes, I. (2017). Evaluating the forensic application of 19 target microRNAs as biomarkers in body fluid and tissue identification. *Forensic Science International: Genetics, 27*, 41–49. https://doi.org/10.1016/j.fsigen.2016.11.012

Sohel, M. H. (2016). Extracellular/circulating microRNAs: Release mechanisms, functions and challenges. *Achievements in the Life Sciences, 10*, 175–186. https://doi.org/10.1016/j.als.2016.11.007

TianYue, M., YiPing, H., & Zheng, W. (2020). Research progress on discrimination of monozygotic twins. *Scientia Sinica Vitae, 50*, 966–972. https://doi.org/10.1360/ssv-2019-0298

Tsokos, M. (2007). Postmortem diagnosis of sepsis. *Forensic Science International, 165*, 155–164. https://doi.org/10.1016/j.forsciint.2006.05.015

Tu, C., Du, T., Shao, C., Liu, Z., Li, L., & Shen, Y. (2018). Evaluating the potential of housekeeping genes, rRNAs, snRNAs, microRNAs and circRNAs as reference genes for the estimation of PMI. *Forensic Science, Medicine and Pathology, 14*, 194–201. https://doi.org/10.1007/s12024-018-9973-y

Tu, C., Du, T., Ye, X., Shao, C., Xie, J., & Shen, Y. (2019). Using miRNAs and circRNAs to estimate PMI in advanced stage. *Legal Medicine, 38*, 51–57. https://doi.org/10.1016/j.legalmed.2019.04.002

van den Berge, M., & Sijen, T. (2017). A male and female RNA marker to infer sex in forensic analysis. *Forensic Science International: Genetics, 26*, 70–76. https://doi.org/10.1016/j.fsigen.2016.10.018

van den Berge, M., Wiskerke, D., Gerretsen, R. R., Tabak, J., & Sijen, T. (2016). DNA and RNA profiling of excavated human remains with varying postmortem intervals. *International Journal of Legal Medicine, 130*, 1471–1480. https://doi.org/10.1007/s00414-016-1438-9

Virkler, K., & Lednev, I. K. (2009). Analysis of body fluids for forensic purposes: From laboratory testing to non-destructive rapid confirmatory identification at a crime scene. *Forensic Science International, 188*, 1–17. https://doi.org/10.1016/j.forsciint.2009.02.013

Wang, H., Mao, J., Li, Y., Luo, H., Wu, J., Liao, M., Liang, W., & Zhang, L. (2013). 5 miRNA expression analyze in post-mortem interval (PMI) within 48 h. *Forensic Science International: Genetics Supplement Series, 4*, e190–e191. https://doi.org/10.1016/j.fsigss.2013.10.098

Wang, Z., Zhang, J., Luo, H., Ye, Y., Yan, J., & Hou, Y. (2013). Screening and confirmation of microRNA markers for forensic body fluid identification. *Forensic Science International: Genetics, 7*, 116–123. https://doi.org/10.1016/j.fsigen.2012.07.006

Wang, Z., Zhang, J., Wei, W., Zhou, D., Luo, H., Chen, X., & Hou, Y. (2015). Identification of saliva using micro RNA biomarkers for forensic purpose. *Journal of Forensic Sciences, 60*, 702–706. https://doi.org/10.1111/1556-4029.12730

Weber, J. A., Baxter, D. H., Zhang, S., Huang, D. Y., How Huang, K., Jen Lee, M., Galas, D. J., & Wang, K. (2010). The microRNA spectrum in 12 body fluids. *Clinical Chemistry, 56*, 1733–1741. https://doi.org/10.1373/clinchem.2010.147405

Williams, T., Soni, S., White, J., Can, G., & Javan, G. T. (2015). Evaluation of DNA degradation using flow cytometry: Promising tool for postmortem interval determination. *The American Journal of Forensic Medicine and Pathology, 36*, 104–110. https://doi.org/10.1097/paf.0000000000000146

Wu, S.-C., Yang, J. C.-S., Rau, C.-S., Chen, Y.-C., Lu, T.-H., Lin, M.-W., Tzeng, S.-L., Wu, Y.-C., Wu, C.-J., & Hsieh, C.-H. (2013). Profiling circulating microRNA expression in experimental sepsis using cecal ligation and puncture. *PLoS One, 8*, e77936. https://doi.org/10.1371/journal.pone.0077936

Xiao, C., Pan, C., Liu, E., He, H., Liu, C., Huang, Y., Yi, S., & Huang, D. (2019). Differences of microRNA expression profiles between monozygotic twins' blood samples. *Forensic Science International: Genetics, 41*, 152–158. https://doi.org/10.1016/j.fsigen.2019.05.003

Yang, X., Wang, J., Guo, S.-L., Fan, K.-J., Li, J., Wang, Y.-L., Teng, Y., & Yang, X. (2011). MiR-21 promotes keratinocyte migration and re-epithelialization during wound healing. *International Journal of Biological Sciences, 7*, 685. https://doi.org/10.7150/ijbs.7.685

Yu, S., Na, J.-Y., Lee, Y.-J., Kim, K.-T., Park, J.-T., & Kim, H.-S. (2015). Forensic application of microRNA-706 as a biomarker for drowning pattern identification. *Forensic Science International, 255*, 96–101. https://doi.org/10.1016/j.forsciint.2015.06.011

Zubakov, D., Boersma, A. W., Choi, Y., van Kuijk, P. F., Wiemer, E. A., & Kayser, M. (2010). MicroRNA markers for forensic body fluid identification obtained from microarray screening and quantitative RT-PCR confirmation. *International Journal of Legal Medicine, 124*, 217–226. https://doi.org/10.1007/s00414-009-0402-3

CHAPTER 19

NGS profiling of water and soil microbial DNA in forensic science

Jillian E. Malbrough and Kelly M. Elkins
Department of Chemistry, Forensic Science Program, Towson University, Towson, MD, United States

NGS approaches for microbial profiling of soil and water

Millions of DNA profiles have been collected and analyzed in support of forensic cases; most of these have been for human identification. However, a DNA profile does not reveal the circumstances or location associated with the transfer of biological material. DNA analysis can be used to identify suspect(s) and link their presence or actions to the victim(s) and the scene of the crime. Pet and other nonhuman DNA, such as that from insects, pollen, and vegetation has been instrumental in solving cases. Microbes have been central to solving forensic cases as they have been used as biological weapons, but more commonly they are spread from crime scenes through contact with individuals, objects, and environments. Microbes are ubiquitous in soil, water, on the human skin, and in the human gut, but samples vary in the species and relative abundance of the microbe populations present due to environmental conditions (Mishra et al., 2023).

The use of microbes in forensic science is still in its infancy. Microbes can include insects, archaea, bacteria, viruses, and fungi (Allwood et al., 2020; Santiago-Rodriguez & Cano, 2016). There are different aspects of microbes that can be of interest in a forensic setting, such as using microbes found in soil or water to help pinpoint where a crime was committed, if a suspect can be associated with a place, circumstances of drowning cases, or to determine the microbial agent in bioterrorism investigations (Budowle et al., 2014; Giampaoli et al., 2014; Oliveira et al., 2018; Wang et al., 2020). The effect of environmental conditions on the DNA profile has the potential to be useful for estimating the time of a crime using the necrobiome (Kumari et al., 2022; Oliveira et al., 2018; Young et al., 2014). Microbes in soil and water can be profiled to solve crimes. Given the soil composition complexity in chemical and biological terms, the inherent degree of genetic information could yield a highly site-specific fingerprint in forensic investigations (Concheri et al., 2011).

In particular cases, the soil particulates collected from victims or suspects can be used in predictive geolocation to predict the location of the crime (Pirrie et al., 2017). If the particulates found on a suspect's shoe or vehicle match those of the scene, the individual can be associated with the scene (Kuiper, 2016; Pirrie et al., 2017). A comparison of soil features can be of great significance in a forensic case. Previously, soil profiling was

Next Generation Sequencing (NGS) Technology in DNA Analysis
ISBN 978-0-323-99144-5, https://doi.org/10.1016/B978-0-323-99144-5.00025-1
© 2024 Elsevier Inc.
All rights reserved.

focused on identifying the different minerals and inorganic compounds in the soil to analyze the color, texture, particle size, and density (Concheri et al., 2011; Sangwan et al., 2020). Morphology, including visual appearance, color, particle size distribution, composition, and texture, can be analyzed using optical microscopy, scanning electron microscopy, X-ray diffraction (XRD), laser ablation-inductively coupled plasma mass spectrometry, transmission electron microscopy, and stable isotope analysis (Pirrie et al., 2017). Giampaoli et al. (2014) compared the use of next-generation sequencing (NGS) to geologic analysis of soil (color, XRD crystalline phases, and mineral abundance) and analyzed bacteria, eukaryotes, and plant material present in the soil to differentiate six soil samples by principle component analysis. In one case, soil evidence and geolocation were used to draw a confession from the suspect (Guo et al., 2021). Soil was found on the suspect's clothing and in the area where the victim's body was discovered (Guo et al., 2021). Analysis methods were able to identify different minerals, diatoms, elements, and pollen spores in comparing the dirt from the victim to the suspect (Guo et al., 2021).

Soil and water differentiation using color and texture can be subjective. DNA can be used to analytically differentiate soil and water samples. Previous soil DNA research utilized denaturing gradient gel electrophoresis, which allowed discrimination using 10 distinct bands (Muyzer et al., 1993). An alternate method suggested for differentiating DNA sequences from samples (including autosomal and mitochondrial DNA) is through the use of NGS, a non-Sanger-based DNA sequencing technology (Yang et al., 2014).

As it has been on the market for almost 20 years, the cost has dropped significantly, enabling more forensic labs to adopt NGS technology for use in forensic cases for a variety of applications (Elkins et al., 2021). Additionally, the high throughput is unparalleled: the sequencing reactions can be performed on multiple samples and target regions simultaneously. Finally, no capillary or gel electrophoresis is needed as the sequences are read directly during sequencing by synthesis steps. NGS has been applied to microbial forensics cases, including detecting and differentiating *Bacillus anthracis* (anthrax) and *Yersinia pestis* (plague) in samples and evaluating flora and fauna in environmental samples (Kuiper, 2016; Yang et al., 2014).

Early research studied the bacterial diversity of a soil sample (Stackebrandt et al., 1993) but did not compare several locations, including crime scenes. NGS can be used to analyze a wider range of microbiological diversity (Giampaoli et al., 2018). In a more recent study using the 18S rRNA gene, soil from 11 environments (including forest, grassland, fields, and an urban park) was compared using NGS on the Roche 454 platform (Lilje et al., 2013). Whole genome amplification (WGA), DNA sequencing, and metagenomic classification were used to differentiate samples from two residential parks that were similar in color (Khodakova et al., 2014). The indoor microbiome was investigated in houses in Germany both spatially and temporally (Weikl et al., 2016). The fungal composition was influenced more by the outdoor vegetation and particulates than the bacterial composition, but both changed over time (Weikl et al., 2016). A more recent study reported that the soil

location is more important than the soil type of the sample (Habtom et al., 2019). The microbiome changes from location to location even if the soil type remains the same (Habtom et al., 2019). With microbial DNA profiling being a new approach in forensic investigations, validation of kits and methods is necessary for their implementation in a forensic laboratory (Kuiper, 2016). There is still much work that needs to be done before microbial forensics is widely used for investigative purposes. In this chapter, we will review the current kits and technology available for use.

The first step in the analysis of microbial DNA is sampling, followed by DNA extraction. The samples must be systematically extracted and stored in a cold place so that the microbe representation in the sample does not change. The extracted DNA is quantified to determine if sufficient samples have been isolated, and a quick polymerase chain reaction (PCR) assay is used to determine if PCR inhibitors such as humic acid have been eliminated in the extraction process and if the samples amplify prior to NGS. Conserved regions, such as the 16S rRNA gene are used for this application. Finally, the samples are sequenced to determine the identity of the microbe(s). The 16S rRNA gene can be targeted as it has nine hypervariable (HV) regions that can be used to identify bacteria by comparison to databases, and the constituent species or phyla can be analyzed by metagenomics. Alternatively, a sample can be sequenced and identified using whole genome sequencing (WGS), which is limited to known species (Laurence et al., 2014).

Extraction methods for obtaining microbial DNA from soil and water

When analyzing any sample whether it be from a human subject, an article of clothing, soil, or even samples in a bioterrorism case, the extraction of the DNA from the sample is the critical first step. Steps, including UV irradiation, must be taken to ensure that there is no contamination (Laurence et al., 2014). Every accredited forensic laboratory has standard operating procedures (SOPs) in place to avoid contamination, and the validation of methods ensures that any instruments and methods being used in the DNA recovery process are reliable and robust.

In soil samples, the first step of the extraction process is the most important. To properly extract the DNA from the soil, the analyst must lyse all of the cells to ensure that DNA is not washed away with the soil debris in subsequent steps. There are two approaches to soil DNA extraction: In indirect extraction, the cells are separated from the soil prior to lysing, whereas the cells are lysed open in the presence of soil debris in direct extraction. When possible, direct extraction is the preferred method as it involves fewer steps, which correlates to fewer chances for contamination or sample loss.

One of the considerations when trying to decide upon an extraction method or kit is that not all available kits have been tested for all soil types. The kit can greatly affect DNA extraction yields and efficiency, and some kits may overrepresent or underrepresent certain microbe species within the sample. Ensuring the kit used is optimal for the soil type is crucial. A list of available soil extraction kits is listed in Table 19.1. The input

Table 19.1 Commercially available kits for soil DNA extraction.

Kit name	Company	Soil Mass	Cell lysis/physical treatment method	Elution volume	Comment(s)
DNeasy PowerSoil Pro Kit	QIAGEN	<0.25 g	Bead-beating (different bead sizes and types available)	50–100 μL	Effective at removing PCR inhibitors from even the most difficult soil types. This is an option to do a 96-well extraction.
DNeasy PowerMax Kit	QIAGEN	<10 g	Bead-beating (different bead sizes and types available)	5 mL	Utilizes their patented Inhibitor Removal Technology (IRT) making it possible to process samples that have in the past proven difficult due to high levels of humic-like substances.
DNeasy UltraClean Microbial Kit	QIAGEN	<0.5 g	Bead-beating (ceramic or glass beads are available), heat, and detergent	50 μL	Can be automated with the option to do 4–96 well plates at a time.
SoilMaster DNA Extraction Kit	Epicenter	0.10 g	Hot-detergent lysis proteinase K	300 μL	Utilizes a chromatography step to remove enzymatic inhibitors known to coextract with the DNA.

Soil DNA Isolation Kit	Norgen Biotek	0.25 g	Bead-beating	50–100 μL	Removes all traces of humic acid and PCR inhibitors with an OSR (organic substance removal) solution.
NucleoSpin Soil Kit	Macherery-Nagel	<0.5 g	Offers two alternative lysis buffers to suit soil sample, bead-beatings (ceramic beads)	30–100 μL	For high molecular weight genomic DNA from microorganisms, the NucleoSpin inhibitor removal column helps to remove inhibitors and humic acid from the sample.
SurePrep Soil DNA Isolation Kit	Fisher BioReagents	0.25 g	Bead-beating	50 μL	Removes all humic acid.
E.Z.N.A. Soil DNA Kit	OMEGA Biotek	0.1–0.25 g	Incubation bead beatings (glass beads)	50–100 μL	Unique inhibitor removal reagent (cHTR reagent) for effective removal of humic acid and other PCR inhibitors from eluted DNA.
FavorPrep Soil DNA Isolation Kit (Mini/Midi)	FAVORGEN Biotech Corp	Mini: 0.25–0.5 g Midi: <10 g	Incubation bead beatings (glass beads) proteinase K	Mini: 50–200 μL Midi: 1–2 mL	Soil in lysis buffer is incubated at 70°C for 10 min.

Continued

Table 19.1 Commercially available kits for soil DNA extraction.—cont'd

Kit name	Company	Soil Mass	Cell lysis/physical treatment method	Elution volume	Comment(s)
FastDNA SPIN Kit	MP Biomedicals	<0.5 g	Cell lysis buffer and the use of the FastPrep instrument to disrupt cell walls	50–100 μL	Delivers the highest DNA yield based on citations.
prepGEM Bacteria	MicroGEM	<0.05 g	Temperature-driven enzymatic lysis	100 μL	Green + buffer is needed.
ZymoBIOMICS DNA Miniprep Kit	ZYMO Research	0.25 g	Bead-beating	100 μL	Uses its BashingBeads to break apart tough microbes.
PureLink Microbiome DNA Purification Kit	Invitrogen	0.2 g	Heat, chemical, and mechanical disruption (bead-beating)	50–200 μL	Typical DNA recovery is 0.5–15 μg from 0.2 g of soil sample.

quantity ranges from 0.05 to 10 g. Most of the kits employ bead-beating steps, while some involve heating and enzymatic digestion steps. Some of the methods can be automated, but most are manual extraction methods. The PowerSoil DNA Isolation Kit is recommended for extracting high-quality DNA from soil samples and removing PCR inhibitors; soil can be a difficult substrate from which to recover DNA (Young et al., 2014).

Quantitation methods for microbial DNA

Human identification methods typically employ a quantitation step to ensure that an appropriate quantity of DNA is input for targeted PCR prior to capillary electrophoresis separation or sequencing. Microbe kits can be used for quantitative PCR (qPCR) quantitation. An example is the Qiagen Microbial DNA qPCR Assay Kit. Alternatively, nonspecific DNA quantification methods, including UV-Vis spectroscopy and fluorescence spectroscopy can be used. qPCR methods can be sensitive to picomolar concentrations, while fluorescence is sensitive to nanomolar concentrations, and UV-Vis spectroscopy is sensitive to micromolar concentrations. The NanoDrop and Qubit instruments are rapid and easy-to-use UV-Vis and fluorescence spectrometers, respectively. The Qubit requires a few preparation steps with consumables that the NanoDrop does not, but it has higher sensitivity.

NGS methods for microbial DNA analysis

NGS technology, including the MiSeq instrument by Illumina, has been developed to be used to analyze multiple loci and samples in a single run at a relatively low per-sample cost with an extremely high depth of coverage for microbes and the ability to detect low enrichment species. Although DNA methods have been applied to soil discrimination in the past, the use of NGS to probe 16S and 18S rRNA gene HV region amplicons for soil microbe profiling can lead to more detail. NGS can be used to analyze the microbiome for forensics (Zhang et al., 2022). NGS can discriminate between samples that have alleles of similar lengths and regions with single nucleotide polymorphisms and can be used with metabarcoding to differentiate samples that are geologically very similar but come from different environments (Fadrosh et al., 2014; Young et al., 2016). Soil biodiversity can be analyzed using environmental DNA (eDNA) using NGS metabarcoding (Frøslev et al., 2022). There are optimized NGS analysis workflows for 16S metagenomics, typically using the V3 and V4 HV regions but also using the full 16S region (Fadrosh et al., 2014; Lu & Salzberg, 2020). The 16S method enables taxonomic classification of the microbial communities in the samples.

Illumina instruments employ sequence-by-synthesis (SBS) chemistry; single bases are detected as they are incorporated into DNA strands in a massively parallel process (Kircher & Kelso, 2010). Fluorescent terminator dyes are imaged as each nucleotide (dNTP) is added and then cleaved to allow incorporation of the next base in a reversible terminating process. Since all four reversible, terminator-bound dNTPs are present, incorporation bias is minimized. Plug-and-play reagent cartridges with radio-frequency identification (RFID) tracking are used to reduce the complexity of use and maintenance (Kircher & Kelso, 2010). After sample preparation (extraction, quantitation, PCR for adding adaptors and indices, and NGS library purification), the length of the sequencing process is approximately 7.5–8 h (4 h for clonal amplification, sequencing, and quality-scored base calling, and 3 h for sequence alignment).

The first step in Illumina sequencing using NGS is library preparation, which results in amplified, tagged, and barcoded PCR products. The first PCR step in library preparation is to amplify and tag the target region. The second round of PCR adds a unique barcode to each amplicon and adapters to bind the Illumina sequencing flow cell (Fadrosh et al., 2014). PhiX control can be used to verify that library prep has been completed correctly. PCR products can be purified using the Agencourt AMPure XP magnetic PCR purification system, quantified using a Qubit fluorescence quantification system, and sized with a QIAxcel DNA fragment sizing system or agarose gel. The purified PCR products are normalized before pooling for simultaneous sequencing of numerous targets and tens of samples in parallel on an Illumina instrument.

NGS data analysis includes taxonomy of soil bacterial species identified, number of read base pairs, total number of reads, median read length, and mean read length. Several bioinformatic software tools are used. Phylogenetic and microbial relationships can be determined using metagenomics (Kirubakaran et al., 2020). A barcoding approach for DNA from plants, microflora, metazoa, and protozoa was developed and applied to forensic soil samples (Giampaoli et al., 2014). A semiautomated approach for NGS metabarcoding for fungal analysis was recently published (Giampaoli et al., 2020). The newly created forensic microbiome database tool collects 16S rRNA data from publicly available data with its associated metadata and can be used to interpret microbiome data for forensic geolocation (Singh et al., 2021).

Application of microbial DNA analysis in geolocation in the United States

Soil studies for forensic geolocation have been conducted in Germany and Australia. We have collected and are analyzing the soil microbiome signatures of soil collected from US states and urban and rural environments. The location of each site has been recorded, and the soil DNA has been extracted using several kits, quantitated using a NanoDrop spectrophotometer, and amplified using PCR with primers targeting the

16S rRNA gene (Klindworth et al., 2013). The samples are undergoing 16S library prep to selectively amplify the V3/V4 region of bacteria and archaea for analysis of variation among the locations and to identify distinguishing species in each location.

Ongoing challenges and conclusion

Microbial DNA profiling using NGS is being applied to a variety of applications, including associating a suspect to a crime scene, geolocation, and identification of bioterror agents (Guo et al., 2021; Haarkötter et al., 2021; Kuiper, 2016). NGS has made microbial profiling less expensive and more accessible. Limitations of microbial DNA profiling from soil and water include the completeness of searchable databases for comparison and species identification, the availability of standardized library prep kits, standardized protocols, analysis toolkits for forensic applications, navigating coextracted contaminants, sequencing prior to precloning, and the capability of the sequencer to perform deep sequencing. Lab SOPs will need to be updated to include microbial profiling and standardized workflows; analysis pipelines, data storage servers, database applications, and reporting will also need to be addressed (Budowle et al., 2014; Jurkevitch et al., 2020; Sjödin et al., 2012).

References

Allwood, J. S., Fierer, N., & Dunn, R. R. (2020). The future of environmental DNA in forensic science. *Applied and Environmental Microbiology, 86*(2), e01504–1519. https://doi.org/10.1128/aem.01504-19.2

Budowle, B., Connell, N. D., Bielecka-Oder, A., Colwell, R. R., Corbett, C. R., Fletcher, J., Forsman, M., Kadavy, D. R., Markotic, A., Morse, S. A., Murch, R. S., Sajantila, A., Schmedes, S. E., Ternus, K. L., Turner, S. D., & Minot, S. (2014). Validation of high throughput sequencing and microbial forensics applications. *Investigative Genetics, 5*, 9. https://doi.org/10.1186/2041-2223-5-9

Concheri, G., Bertoldi, D., Polone, E., Otto, S., Larcher, R., & Squartini, A. (2011). Chemical elemental distribution and soil DNA fingerprints provide the critical evidence in murder case investigation. *PLoS One, 6*(6), e20222. https://doi.org/10.1371/journal.pone.0020222

Elkins, K. M., Berry, H. E., & Reese, K. R. (2021). Applications of NGS in DNA analysis. In *Handbook of DNA Profiling* (pp. 1–18). https://doi.org/10.1007/978-981-15-9364-2_19-1

Fadrosh, D. W., Ma, B., Gajer, P., Sengamalay, N., Ott, S., Brotman, R. M., & Ravel, J. (2014). An improved dual-indexing approach for multiplexed 16S rRNA gene sequencing on the Illumina MiSeq platform. *Microbiome, 2*(1), 6. https://doi.org/10.1186/2049-2618-2-6

Frøslev, T. G., Ejrnæs, R., Hansen, A. J., Bruun, H. H., Nielsen, I. B., Ekelund, F., et al. (2022). Treated like dirt: Robust forensic and ecological inferences from soil eDNA after challenging sample storage. *Environmental DNA, 5*(1), 158–174. https://doi.org/10.1002/edn.3.367

Giampaoli, S., Alessandrini, F., Frajese, G. V., Guglielmi, G., Tagliabracci, A., & Berti, A. (November 2018). Environmental microbiology: Perspectives for legal and occupational medicine. *Legal Medicine, 35*, 34–43. https://doi.org/10.1016/j.legalmed.2018.09.014

Giampaoli, S., Berti, A., Di Maggio, R. M., Pilli, E., Valentini, A., Valeriani, F., Gianfranceschi, G., Barni, F., Ripani, L., & Romano Spica, V. (2014). The environmental biological signature: NGS profiling for forensic comparison of soils. *Forensic Science International, 240*, 41–47. https://doi.org/10.1016/j.forsciint.2014.02.028

Giampaoli, S., De Vittori, E., Frajese, G. V., Paytuví, A., Sanseverino, W., Anselmo, A., Barni, F., & Berti, A. (January 2020). A semi-automated protocol for NGS metabarcoding and fungal analysis

in forensic. *Forensic Science International, 306*, 110052. https://doi.org/10.1016/j.forsciint.2019.110052

Guo, H., Yao, Y., Li, Y., Wang, P., Hu, C., Yuan, M., Mei, H., & Zhu, J. (2021). A case study in forensic soil comparison. *Journal of Forensic Sciences, 67*(2), 766−774. https://doi.org/10.1111/1556-4029.14921

Haarkötter, C., Saiz, M., Gálvez, X., Medina-Lozano, M. I., Álvarez, J. C., & Lorente, J. A. (2021). Usefulness of microbiome for forensic geolocation: A review. *Life, 11*(12), 1322. https://doi.org/10.3390/life11121322.5

Habtom, H., Pasternak, Z., Matan, O., Azulay, C., Gafny, R., & Jurkevitch, E. (2019). Applying microbial biogeography in soil forensics. *Forensic Science International: Genetics, 38*, 195−203. https://doi.org/10.1016/j.fsigen.2018.11.010

Jurkevitch, E., & Pasternak, Z. (2020). A walk on the Dirt: Soil microbial forensics from ecological theory to the crime lab. *FEMS Microbiology Reviews, 45*(2), fuaa053. https://doi.org/10.1093/femsre/fuaa053

Khodakova, A. S., Smith, R. J., Burgoyne, L., Abarno, D., & Linacre, A. (2014). Random whole metagenomic sequencing for forensic discrimination of soils. *PLoS One, 9*(8), e104996. https://doi.org/10.1371/journal.pone.0104996.6

Kircher, M., & Kelso, J. (2010). High-throughput DNA sequencing-concepts and limitations. *BioEssays, 2*(6), 524−536. https://doi.org/10.1002/bies.200900181

Kirubakaran, R., ArulJothi, K. N., Revathi, S., Shameem, N., & Parray, J. A. (September 2020). Emerging priorities for microbial metagenome research. *Bioresource Technology Reports, 11*, 100485. https://doi.org/10.1016/j.biteb.2020.100485

Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M., & Glöckner, F. O. (2013). Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Research, 41*(1), e1. https://doi.org/10.1093/nar/gks808

Kuiper, I. (2016). Microbial forensics: Next-generation sequencing as catalyst: The use of new sequencing technologies to analyze whole microbial communities could become a powerful tool for forensic and criminal investigations. *EMBO Reports, 17*, 1085−1087. https://doi.org/10.15252/embr.201642794

Kumari, P., Prakash, P., Yadav, S., & Saran, V. (2022). Microbiome analysis: An emerging forensic investigative tool. *Forensic Science International, 340*, 111462. https://doi.org/10.1016/j.forsciint.2022.111462.7

Laurence, M., Hatzis, C., & Brash, D. E. (2014). Common contaminants in next-generation sequencing that hinder discovery of low-abundance microbes. *PLoS One, 9*, e97876. https://doi.org/10.1371/journal.pone.0097876

Lilje, L., Lillsaar, T., Rätsep, R., Simm, J., & Aaspõllu, A. (2013). Soil sample metagenome NGS data management for forensic investigation. *Forensic Science International: Genetics Supplement Series, 4*, e35−e36. https://doi.org/10.1016/j.fsigss.2013.10.017

Lu, J., & Salzberg, S. L. (2020). Ultrafast and accurate 16S rRNA microbial community analysis using Kraken 2. *Microbiome, 8*, 124. https://doi.org/10.1186/s40168-020-00900-2

Mishra, A., Singh, L., & Singh, D. (2023). Unboxing the black box—one step forward to understand the soil microbiome: A systematic review. *Microbial Ecology, 85*(2), 669−883. https://doi.org/10.1007/s00248-022-01962-5

Muyzer, G., DeWaal, E. C., & Uitterlinden, A. (1993). Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction -amplified genes coding for 16S rRNA. *Applied and Environmental Microbiology, 59*, 695−700.

Oliveira, M., & Amorim, A. (2018). Microbial Forensics: New breakthroughs and future prospects. *Applied Microbiology and Biotechnology, 102*(24), 10377−10391. https://doi.org/10.1007/s00253-018-9414-6

Pirrie, D., Dawson, L., & Graham, G. (2017). Predictive geolocation: Forensic soil analysis for provenance determination. *Episodes, 40*(2), 141−147. https://doi.org/10.18814/epiiugs/2017/v40i2/017016.8

Sangwan, P., Nain, T., Singal, K., Hooda, N., & Sharma, N. (2020). Soil as a tool of revelation in forensic science: A review. *Analytical Methods, 12*(43), 5150−5159. https://doi.org/10.1039/d0ay01634a

Santiago-Rodriguez, T. M., & Cano, R. J. (2016). Soil microbial forensics. *Environmental Microbiogy, 4*(4), 143−171. https://doi.org/10.1128/microbiolspec.emf-0007-2015

Singh, H., Clarke, T., Brinkac, L., Greco, C., & Nelson, K. E. (2021). Forensic microbiome database: A tool for forensic geolocation meta-analysis using publicly available 16S rRNA microbiome sequencing. *Frontiers in Microbiology, 12*, 644861. https://doi.org/10.3389/fmicb.2021.644861.9

Sjödin, A., Broman, T., Melefors, O., Andersson, G., Rasmusson, B., Knutsson, R., & Forsman, M. (2012). The need for high-quality whole-genome sequence databases in microbial forensics. *Biosecurity and Bioterrorism, 11*, S78–S86. https://doi.org/10.1089/bsp.2013.0007

Stackebrandt, E., Liesack, W., & Goebel, B. M. (1993). Bacterial diversity in a soil sample from a subtropical Australian environment as determined by 16S rDNA analysis. *The FASEB Journal, 7*, 232–236.

Wang, L.-L., Zhang, F.-Y., Dong, W.-W., Wang, C.-L., Liang, X.-Y., Suo, L.-L., Cheng, J., Zhang, M., Guo, X.-S., Jiang, P.-H., Guan, D.-W., & Zhao, R. (2020). A novel approach for the forensic diagnosis of drowning by microbiological analysis with next-generation sequencing and unweighted UniFrac-based PCoA. *International Journal of Legal Medicine, 134*(6), 2149–2159. https://doi.org/10.1007/s00414-020-02358-1

Weikl, F., Tischer, C., Probst, A. J., Heinrich, J., Markevych, I., Jochner, S., & Pritsch, K. (2016). Fungal and bacterial communities in indoor dust follow different environmental determinants. *PLoS One, 11*(4), e0154131. https://doi.org/10.1371/journal.pone.0154131

Yang, Y., Xie, B., & Yan, J. (2014). Application of next-generation sequencing technology in forensic science. *Genomics, Proteomics & Bioinformatics, 12*, 190–197. https://doi.org/10.1016/j.gpb.2014.09.001

Young, J. M., Austin, J. J., & Weyrich, L. S. (2016). Soil DNA metabarcoding and high-throughput sequencing as a forensic tool: Considerations, potential limitations and recommendations. *FEMS Microbiology Ecology, 93*(2), fiw207. https://doi.org/10.1093/femsec/fiw207

Young, J. M., Rawlence, N. J., Weyrich, L. S., & Cooper, A. (2014). Limitations and recommendations for successful DNA extraction from forensic soil samples: A review. *Science & Justice, 54*(3), 238–244. https://doi.org/10.1016/j.scijus.2014.02.006.10

Zhang, J., Liu, W., Simayijiang, H., Hu, P., & Yan, J. (2022). Application of microbiome in forensics. *Genomics, Proteomics & Bioinformatics*. https://doi.org/10.1016/j.gpb.2022.07.007

This page intentionally left blank

CHAPTER 20

Applications of NGS in analysis of challenging samples

Safia A. Messaoudi
Department of Forensic Science, College of Criminal Justice, Naif Arab University for Security Sciences, Riyadh, Saudi Arabia

Introduction

DNA typing from biological material recovered from a crime scene and human remains for human identification has proven to be a difficult task for forensic geneticists due to various factors that influence the success or failure of testing. The biological evidence that forensic genetics experts get for analysis is not always the most suitable type of sample for forensic DNA analysis, especially in terms of quantities (Whitaker et al., 2001; Chung et al., 2004; Dixon et al., 2006; Gill et al., 2007). The investigation of ancient DNA also presented many challenges in obtaining enough template DNA quantity to test (Llamas et al., 2017).

Conventional techniques requiring the typing of short tandem repeat (STR) markers using capillary electrophoresis (CE) frequently fail in these cases. A possible explanation for this might be the length of the amplified fragment or the signal of imbalance observed between the short and long tandem repeats (Chung et al., 2004; Dixon et al., 2006).

DNA typing through STR analysis has become one of the most well-known tools and the gold standard for human identification (Jordan & Mills, 2021). A major advantage of this approach is the standardization of loci utilized by all laboratories and the extremely large searchable database of genetic profiles. However, some limitations and challenges are faced during the DNA typing of highly degraded or low template DNA samples. Alternative approaches have been developed to overcome these limitations in an attempt to improve the chances of acquiring a genetic profile from such challenging samples, particularly useful in studying mitochondrial DNA (mtDNA), single nucleotide polymorphisms (SNPs) (Brenner & Weir, 2003), insertion/deletion (indels) polymorphisms (Pereira et al., 2009), and mini-STRs (Holland et al., 2003).

CE-based methods were and are the standard method used in forensic analysis and were selected for their reliability and validity. However, the situation necessitates the search for advanced ways to aid the analysis of the challenging DNA evidence.

The introduction of novel techniques and technologies into the criminal justice system is slow, but new methodologies are being developed that enable the analysis of these challenging samples and that can generate intelligence about the donor of a biological sample (McCord et al., 2019; Butler & Willis, 2020). Massively parallel sequencing

Next Generation Sequencing (NGS) Technology in DNA Analysis
ISBN 978-0-323-99144-5, https://doi.org/10.1016/B978-0-323-99144-5.00011-1
© 2024 Elsevier Inc.
All rights reserved.

(MPS) or next-generation sequencing (NGS) could overcome the limitations of the previous technologies used to analyze DNA evidence and have brought numerous solutions to assess degraded DNA evidence, low-template DNA, DNA mixtures, and ancient DNA.

This chapter highlights the types of challenging DNA samples, the previous technology used to analyze them, the important application of NGS, and the available gene panels that have been utilized for the abovementioned evidence.

Application of next-generation sequencing technology in forensic science

The use of NGS for forensic applications has expanded rapidly in the last few years, and its incorporation into forensic workflows is becoming a realistic choice (Kayser & Parson, 2018; de Knijff, 2019). NGS has been chosen mainly due to the continuous decrease in costs, the implementation of initial PCR-amplification of a set of target markers followed by MPS of the resulting amplicons, and the development of bioinformatics tools that enable the analysis of the large volume of complex data produced (Alonso et al., 2018; Bruijns et al., 2018).

One of the major advantages of NGS methods is the ability to incorporate a large number of different marker types into a single assay, as well as shorter amplicons compared to conventional STR profiling methods, improving the results, power of discrimination, and robustness to analyze damaged and limited DNA input samples (Shih et al., 2018; Zhang et al., 2018; Borsuk et al., 2018; Amankwaa & McCartney, 2021). Another significant benefit is the capability to detect nucleotide sequence variation in the targeted markers, such as variants in STR repeat regions and flanking sequences, which is useful in complex mixture profile interpretation and therefore enables the discrimination of alleles that would be indistinguishable using CE-STR-based typing (Pereira et al., 2018; Novroski et al., 2019). These developments explain how NGS technology continues to impact forensic science positively.

Despite the success of the NGS technologies, they are not yet widely used in casework in other areas of forensic analysis that are indispensable tools for the forensic community such as RNA sequencing, epigenetics, and forensic DNA phenotyping (Yang et al., 2014).

Panels developed for the forensic samples using the NGS technologies

Several NGS-based sequencing platforms, also called second-generation sequencing, have been developed by Roche Life Sciences, Thermo Fisher Scientific, Illumina, and Applied Biosystems, including pyrosequencing, Ion Torrent technology, the Illumina/Solexa platform, and SOLiD (Diegoli et al., 2012). Table 20.1.

Table 20.1 Panels developed for the forensic samples using the NGS technologies.

Platform	Application	Kit	Markers
Ion gene studio S5 system, Ion PGM system, Ion S5 XL system, Ion OneTouch 2 system, Ion Chef system	Analyze highly degraded or trace DNA Generate more investigative leads in development Analyze familial and kinship cases Analyze missing persons and remains Analyze DNA mixtures with more efficiency	Precision ID mtDNA Whole Genome Panel	Analyze the entire human mitochondrial genome (16,569 bp).
		Precision ID mtDNA Control Region Panel	It is a 2-pool multiplex assay that targets the 1.2 kb control region of the human mitochondrial genome, which encompasses HV-I, II, and III.
		Precision ID GlobalFiler NGS STR Panels	The panel targets 35 markers comprised of 21 CODIS STR markers, 9 additional multiallelic STR markers, and 4 sex determination markers.
		Precision ID Ancestry Panel	The panel includes 165 autosomal markers (SNPs).
		Precision ID Identity Panel	The panel includes 34 upper Y-clade SNPs and 90 autosomal SNPs.
		Ion AmpliSeq VISAGE-Basic Tool Research Panel	Panel consists of 153 SNPs, allowing prediction of EVCs and biogeographical ancestry and incorporating 41 SNPs of the HIrisPlex-S assay1 and 115 BGA markers chosen for differentiation of seven continental populations.
		Ion AmpliSeq DNA Phenotyping Panel	This panel targets 23 SNPs and one indel from the following genes: MC1R, SLC45A2, ASIP/PIGU, EXOC2, HERC2, IRF4, KITLG, OCA2, TYR, SLC24A4, and TRYP1.

Continued

Table 20.1 Panels developed for the forensic samples using the NGS technologies.—cont'd

Platform	Application	Kit	Markers
		Ion AmpliSeq PhenoTrivium Panel	Contains ancestry, DNA phenotyping, and male lineage markers for a total of 200 autosomal SNPs and 120 Y chromosomal SNPs.
		Ion AmpliSeq HID Y-SNP Research Panel v1	The panel enables analyses of 859 Y-SNPs to infer 640 Y haplogroups.
		Ion AmpliSeq MH-74 Plex Research Panel	The Panel is a 157−325 bp assay covering 74 microhaplotypes (230 SNPs) selected from a set of 130 microhaplotypes.
Illumina (MiSeq)	Forensic DNA testing for criminal casework	ForenSeq DNA signature prep Kit	230 genetic markers: 27 global autosomal STRs, 24 Y-STRs, 7 X-STRs, 95 identity SNPs, 22 phenotypic SNPs, and 56 biogeographical ancestry SNPs.
	mtDNA analysis to identity missing persons	Verogen ForenSeq mtDNA Control Region Kit	Analyze the mtDNA D-loop hypervariable region. For difficult samples and amplify four overlapping fragments spanning the hypervariable regions.
		Verogen ForenSeq mtDNA Whole Genome Kit	For clean, intact samples and amplify two long fragments spanning the entire human mitochondrial genome (16,569 bp).
	Disaster victim identification	Combining the ForenSeq DNA signatureand the Nextera XT DNA Library prep Kits	

		Verogen ForenSeq MainstAY Kit	For targeted sequencing of established autosomal and Y-STR markers. Simultaneous analysis of 52 genetic markers. *ForenSeq MainstAY Kit:* 27 global autosomal STRs including, CODIS and the European standard set of 25 Y-STRs, including the minimum set for the Y Haplotype Reference Database (YHRD). *ForenSeq MainstAY SE Kit:* Includes all markers in ForenSeq MainstAY Kit +SE33.
	For long range kinship and genetic genealogy use. Dedicated QIAseq panels for human identity investigations.	Verogen ForenSeq K intelligence Kit QIAseq investigator Panels (https://www.qiagen.com/us/products/human-id-and-forensics/nextgeneration-sequencing/qiaseq-investigator-panels/)	Simultaneous analysis of 10,230 SNPs.

Qiagen. (2022). "Next-Generation Sequencing for HID." Retrieved January 10, 2023, 2023, from https://www.qiagen.com/us/product-categories/human-id-and-forensics/nextgeneration-sequencing. Thermo Fisher Scientific. (2022). "Precision ID NGS System for Human Identification: Targeted Sequencing for Forensic DNA Analysis." from https://www.thermofisher.com/sa/en/home/industrial/forensics/human-identification/forensic-dna-analysis/dna-analysis/next-generation-sequencing-ngs-forensics.html.

Furthermore, continuous optimization has resulted in innovative third- and fourth-generation platforms such as PacBio's single-molecule real-time sequencing, nanopore sequencing, etc. Since 2012, these available instruments have significantly increased the number of sequenced genomes and published genome-based studies (Barba, n.d., pp. 149–169; Braslavsky et al., 2003; Harris et al., 2008).

Similarly, a wide range of validated and/or evaluated kits targeting various forensically relevant markers are currently available for use on each platform (Wang et al., 2017; Alonso et al., 2018; Ballard et al., 2020). These are especially: the Precision ID GlobalFiler NGS STR Panel and Precision ID Identity/Ancestry Panels, using the Ion S5 system to sequence STR and SNP markers (Meiklejohn & Robertson, 2017; Montano et al., 2018); Promega's PowerSeq system that sequences autosomal and Y-STRs using the Illumina MiSeq platform (Guo et al., 2017) and the Verogen ForenSeq DNA Signature Prep Kit, designed to work with the MiSeq FGx Forensic Genomics System, which uses Illumina technology to sequence autosomal STRs, Y-/X-STRs, and identity SNPs, and can be expanded to include phenotype- and ancestry-informative SNPs (Moreno et al., 2018; Pereira et al., 2018).

Several Kits have also been developed to target part or all of the mitochondrial genome, such as the Precision ID mtDNA Whole Genome Panel that amplifies the entire mtDNA genome using the Ion AmpliSeq technology, Nextera XT library preparation, and Illumina sequencing on the MiSeq FGx system, as well as the Verogen ForenSeq mtDNA Control Region Kit also using the MiSeq FGx Sequencing System (Peck et al., 2018; Pereira et al., 2018; Holt et al., 2019).

DNA damage and degradation

Human biological specimens submitted for forensic DNA analysis are rarely unaffected by exposure to their environment, and consequently, small amounts of DNA, which are generally highly damaged are frequently extracted from this kind of sample. Several factors such as heat, humidity, pH of the aqueous environment, ultraviolet (UV) light, chemical agents, and enzymatic attack can contribute to the damage that samples collected from crime scenes, mass disasters, and mass graves' locations can cause, and it is difficult to determine the extent of damage when the DNA of interest is present in a sample containing DNA from multiple sources (Hall & Ballantyne, 2004).

Commonly, the DNA is damaged through various endogenous processes such as oxidation, hydrolysis, pyrimidine dimers, cytosine deamination, DNA-DNA cross-linking, and DNA-Protein cross-linking (Diegoli et al., 2012). The most common forms in forensic samples are oxidation, hydrolysis, and pyrimidine dimers. The oxidative damage is mediated by free radicals and leads to the creation of modified oxidative base products that prevent the DNA replication process during PCR by inhibiting the standard

Taq DNA polymerase to bypass the lesions due to the lack of 3′ exonuclease activity (Evans et al., 1993; Lindahl, 1993).

It has been reported that hydrolysis and oxidation still take place even in the solid-state DNA and can only be avoided if the biological evidence is protected from atmospheric water and oxygen, which is actually difficult, especially since these conditions are commonly encountered in cases of natural and mass disasters (flooding, tsunamis, waterlogged burial sites, etc.) (Colotte et al., 2011).

Additionally, photochemical reactions induced by UV radiation generated the formation of covalent linkages by localized reactions on the C=C, or T=T double bonds, which stall DNA polymerase and arrest replication during PCR (Goodsell, 2001). UVC radiation-induced DNA damage makes forensic DNA analysis partially or completely nonfunctional by affecting PCR efficiency and, subsequently, DNA profile identification (Gršković et al., 2013).

The passage of time and the improper conditions of storage and processing of biological materials also amplify the risk of DNA degradation. This could be due to the obstinate microbial fauna and flora growth, which are also common contaminants in skeletal material recovered from a crime scene or a mass disaster (Van der Schans, 1978). Microorganism growth has been shown to be a significant cause of DNA damage leading to the nontypeability of forensic samples by secreting digestive enzymes that induce DNA damage via reactive oxygen species and double-stranded breaks that can increase by 25-fold (Ballantyne, 2006). The microbial degradation can also lead to inappropriate or no PCR amplification as well as the detection of heterozygous peak imbalances (Dash & Das, 2018).

While the causes of DNA degradation differ, the final result remains the same: the fragmentation of the double helix into small pieces, leading to a significant loss of genetic information inadequate for PCR amplification and resulting in incomplete or no DNA profiles (Bender et al., 2004; Ballantyne et al., 2007; Fondevila et al., 2008).

Several pieces of research have revealed that PCR inhibition is the most common cause of PCR failure and remains a big challenge in forensic DNA analysis of biological remains recovered from the environment (Alaeddini, 2012). These inhibitors, which are found in a variety of biological materials (hemoglobin, myoglobin, bile salts), environmental samples (fulvic acids, humic acids, humic material Metal ions), and food (calcium ions, Proteases) (Schrader et al., 2012), may negatively affect cell lysis during extraction, primers, or reduce the polymerase activity (Hedman & Rådström, 2013).

Overall, regardless of the source, the DNA damage may significantly interfere with PCR amplification and therefore decrease the genotyping success, especially when the DNA is isolated from a piece of biological evidence discovered at a crime scene that is not under homeostatic control, resulting in the accumulation of mutations over time and potentially allele drop-outs in DNA profiles (Hall & Ballantyne, 2004).

Challenging DNA evidence

DNA serves various functions in the human body, including coding for proteins. DNA also facilitates inheritance and provides a means to explain disparate aspects of life as well as its various processes (Travers & Muskhelishvili, 2015). In addition, DNA has proven its effectiveness as a powerful tool in criminal investigations (Amankwaa & McCartney, 2021). However, forensic DNA laboratories encounter many challenges concerning DNA evidence due to degradation, the presence of PCR inhibitors, low-template DNA samples, and mixture profile interpretation (Butler, 2015). Forensic geneticists who are frequently confronted with low- or poor-quality samples are constrained to provide partial genetic profiles with allele and/or locus drop-out, equivocal results, or, worse, no results. Until now, to resolve these issues, the most effective approach has been to increase PCR efficiency by generating shorter amplicons (Constantinescu et al., 2012).

Highly degraded biological evidence significantly reduced the quality of the DNA polymer, making effective testing of the sample difficult, especially with the standard nuclear DNA testing protocols, which seem ineffective with severely degraded samples and necessitate further methods (Burger et al., 1999).

When it comes to DNA extraction from biological evidence, ancient DNA (aDNA) also presents many challenges, necessitating the use of advanced methods. Contamination is one of the issues that scientists face with this type of DNA, which leads to exogenous genetic materials (Llamas et al., 2017). The challenges arise because the available DNA in the fossil is generally scarce and fragmented due to the environmental microorganisms as well as the people who handled the fossils (Slatkin & Racimo, 2016), and any action that contaminates the aDNA aggravates the challenge of obtaining reliable evidence (Machado & Granja, 2020, pp. 45–56).

Low-template DNA, or low copy number DNA, is another challenge that forensic investigators face during the analysis of DNA evidence due to the collection of shreds of samples containing small amounts of DNA (Tan, 2010, pp. 1–45). Traditionally, the extraction of DNA from this type of evidence is not successful, and the related genetic profiles are sometimes insignificant, making them inadequate for evaluations and, in most cases, leading to the nullification of evidence and therefore obstructing those investigations (Mateen et al., 2020).

Another major challenge is the interpretation of DNA mixture profiles containing contributions from multiple donors due to the low sensitivity of STR loci to allow the detection of the minor donor in the presence of a major donor, not only because of the potential number of alleles shown in the profile but also because of the low-level profiles with common allele drop-out/drop-in and heterozygous imbalance (Gill et al., 2015).

The existence of additional donors from the major or minor DNA fraction increases the complexity of the mixtures, which has a significant impact on the downstream interpretation of STR profiles. Recent advances in STR profiling techniques have facilitated the investigation and recovery of DNA mixtures, not only from samples where mixtures might be expected (e.g., sexual offense samples), but also from low-quality/quantity samples recovered from handled items that often produce complex mixtures with large numbers of contributors (Kelly et al., 2012; Cavanaugh & Bathrick, 2018). Therefore, and due to the increasing complexity of DNA mixture profiles, a significant deviation from relatively simple genotyping methods for determining whether an individual could be excluded as a potential contributor to a mixture to the use of likelihood ratio tests has facilitated the estimation of the most likely genotype combinations of contributors to a mixture (Kelly et al., 2014; Bieber et al., 2016).

All these challenges have led to the development of new methods, especially of DNA extraction and PCR approaches and kits, as well as methods of interpretation using probabilistic frameworks, incorporating probabilities of allele drop-out and drop-in, modeled from validation and empirical data (Bright et al., 2017; Coble & Bright, 2019). NGS is a well-established approach that could help overcome some of the STR mixture constraints and improve the analytical capabilities of challenging mixture samples (Oldoni & Podini, 2019).

Previous types of technologies used to analyze challenging DNA

CE remains the most well-known tool for assessing the separation of DNA in forensic fields by amplicon length. However, as mentioned previously, the small amounts of DNA are a serious challenge that forensic scientists grapple with when collecting evidence from crime or accident scenes. Degraded human remains also lead to low-template DNA, necessitating the need to adopt alternative methods for DNA analysis. Different methods have been proposed to assess the challenging forensic DNA samples, which are:

MiniSTRs

One of the previously used approaches to analyzing challenging DNA samples is miniSTR analysis. This approach resolves the issues of DNA quality and quantity that are often a problem when handling degraded samples by employing reduced-size STR amplicons for the analysis of low-template DNA (Butler & Willis, 2020). Through miniSTRs, scientists can recover the needed information from the small samples in their possession (Mautner et al., 2017).

PCR is the mechanism by which miniSTR operates to amplify DNA regions that are highly polymorphic and repetitive (Jordan & Mills, 2021). By designing primer binding

sites that are closer to the target regions, which reduce the amplicon size to 80–150 bp, miniSTR analysis can be used for degraded DNA samples (Butler et al., 2003; Dixon et al., 2006; Opel et al., 2007).

However, there are certain drawbacks associated with the use of miniSTRs, especially that while bringing the primers closer to the repeat region reduces overall amplicon size, loci with the widest range of alleles, such as D21S11 and FGA, that remain have large allele size differences, leading to a difference in PCR performance between the extreme repeat numbers (Gill, 2001).

Mitochondrial DNA typing

The fact that mtDNA does not change from the mother to their offspring explains its wide application in forensic investigations. This type of inheritance makes this small genome essential to the determination of a person's ancestry as well as the analysis of degraded or low-copy-number DNA (Mishmar et al., 2019). The analysis of mtDNA sequence polymorphisms has proved useful when nuclear DNA is too degraded since mtDNA is more likely to be present in forensic biological samples in high copy numbers. This is evident in the cases of telogen hair roots, hair shafts, bones, teeth, and touch DNA (Budowle et al., 1999; Parson et al., 2014).

The effectiveness of mtDNA and its vitality for DNA evidence analysis have been reported in its resistance to environmental insults. There have been several investigations analyzing the whole mtDNA rather than just the hypervariable polymorphic regions HVI/HVII to increase the power of discrimination of the mtDNA and to resolve the problem of common sequences found in the HVI/HVII regions shared by populations of the same origin (Templeton et al., 2013; Zavala et al., 2019; Budowle et al., 2004; Álvarez-Iglesias et al., 2007).

Until recently, the only approach utilized as an alternative to standard STR genotyping for highly degraded DNA was mtDNA sequencing following the standard protocols. However, there are certain difficulties with the use of mtDNA sequencing, mainly because it's a laborious, time-consuming, and costly approach for the operational laboratory. Therefore, only a portion of forensic DNA laboratories have implemented mtDNA analysis (Budowle & Van Daal, 2008). Another advantage of using mtDNA typing in casework is that the observed power of discrimination is not as high as that provided by the STR loci, a consequence of the matrilineal inheritance rules in certain relationship comparisons such as father and daughter as well as the urgent need of large-scale haplotype databases to interpret correctly the significance of mtDNA variation (Álvarez-Iglesias et al., 2007). Another problem with this approach is that it fails when mtDNA suffers from heteroplasmy, where more than one mtDNA genome type can be present in individuals (Mechta et al., 2017).

Single nucleotide polymorphism typing

SNPs are single-base variations, substitutions, insertions, or deletions at a unique physical location. SNPs are highly abundant, estimated to occur at 1 in every 1000 bases in the human genome, and represent the most common class of human polymorphism (Sachidanandam et al., 2001; Venter et al., 2001). Nevertheless, there is an abundance of genetic information that can be exploited since approximately 85% of human variation is derived from SNPs (Cooper et al., 1985; Wang et al., 1998; Holden, 2002).

According to previous reports, SNP markers are valuable for DNA typing of damaged samples by improving the quantity of genetic information obtained from challenging forensic samples (Gill, 2001; Budowle et al., 2004; Sobrino et al., 2005). It has been reported that SNPs are powerful on degraded and highly degraded samples and can be analyzed with rapid-high-throughput platforms using short amplicons, which is preferable for the success of amplification in most cases due to the small size of the amplified product (Dixon et al., 2006; Butler et al., 2007; Budowle & Van Daal, 2008). In this case, it is possible to optimize the PCR reaction to amplify the SNP markers with designed primers, immediately adjacent to targeted single base variations, required to amplify a 60–80 bp fragment (Budowle, 2004; Budowle et al., 2004; Divne & Allen, 2005; Sobrino et al., 2005; Dixon et al., 2006).

In addition, SNP analysis would take a substantial reduction in cost, increased throughput, and enhanced capabilities to resolve mixed samples for the forensic community to replace the STR loci (Gill et al., 2004; Sobrino et al., 2005; Butler et al., 2007; Oldoni & Podini, 2019). Several studies reported the effectiveness of SNP variants, particularly using high-throughput technologies, which have become increasingly important for the successful implementation of large criminal DNA databases (Divne & Allen, 2005; Børsting & Morling, 2016; Guo et al., 2017; Shih et al., 2018; Zavala et al., 2019; Haddrill, 2021).

Indel polymorphisms are among the most suitable technologies for DNA testing that can help scientists overcome any issues arising from DNA's low-template DNA or degradation due to the small-sized amplicons that make indels the most appropriate for genotyping DNA that has suffered high degradation (Pereira et al., 2009; Zidkova et al., 2013; Klein et al., 2015). Indels were also chosen to be suitable for paternity cases due to their low rate of mutation compared with STRs (Klein et al., 2015). The indels-based method involves the identification of a specific mutation that provides diallelic markers (Pereira et al., 2009; Klein et al., 2015). One of the most well-known commercial kits for assessing indels polymorphisms is the Investigator DIPlex kit (Qiagen), which enables the forensic community to utilize this method in their investigations and was likely effective in challenging DNA analysis (Jordan & Mills, 2021; Klein et al., 2015). The kit can cover 30 Indel loci that are distributed in more than 9 autosomal chromosomes (Zidkova et al., 2013; Klein et al., 2015).

Due to their biallelic profile and the well-established STR loci databases, SNPs are unlikely to become the primary forensic markers and are considered not as informative per locus as the forensically selected STR loci. Another major problem with the SNP-based methods is that a panel of at least 50−100 autosomal SNP loci would be required to achieve the same power of discrimination that the existing STR multiplex systems can provide (Chakraborty et al., 1999).

Application of NGS in challenging DNA evidence

CE is the primary method used in forensic DNA laboratories around the world for separating and detecting STR alleles (Butler, 2012). However, the previously reported limitations to assessing effectively the challenging DNA samples necessitate further developments in advanced tools to overcome those problems by targeting the entire genome. In particular, the analysis of standard forensic STR markers using MPS revealed several advantages over conventional CE, including (1) An increased number of loci that can be analyzed simultaneously, (2) A higher discrimination power due to increased STR allele sequence diversity and the greater number of loci, and (3) Shorter amplicons for a more effective analysis of degraded and/or low-quantity forensic biological evidence (Alonso et al., 2018).

Given these benefits, NGS is a promising adjunct method for forensic casework in spite of the recommendation of the International Society for Forensic Genetics that the use of NGS in casework necessitates the availability of global datasets designed to enrich the allele frequencies of sequence-based STR genotypes (Parson et al., 2016; Phillips et al., 2018a,b).

Degraded samples and low-template DNA evidence.

STRs are the markers used for routine human identity testing in the forensic community. However, despite their high discrimination and robustness, STR assays frequently fail when genotyping extremely degraded samples (Butler et al., 2003; Ballantyne et al., 2007; Bogas et al., 2009; Dash & Das, 2018; Haddrill, 2021). The NGS approach has become one of the more practical ways to address the issues that forensic scientists and other DNA laboratory specialists have suffered when handling limited DNA evidence due to its discrimination power, which can be increased through the simultaneous analysis of multiple, diverse genetic markers, including not only STRs but also SNPs and mtDNA (Zavala et al., 2019). It has been reported that NGS technology has significantly improved results for low-level and degraded DNA samples as a result of the shorter amplicons compared with standard STR profiling (Haddrill, 2021). Another major advantage of NGS technology is its capacity to characterize single molecules, and therefore it can deliver results with much lower concentrations of input DNA (Prosser et al., 2016).

Previously, forensic scientists required considerable DNA input to analyze and generate evidence that could be used in a given case. However, contamination and degradation could not allow these professionals to obtain those large amounts, derailing and reducing the quality of their work. But fortunately, the discovery of NGS technology has reduced the DNA sample input that forensic scientists require to prepare NGS libraries for sequencing to obtain maximum information (Alvarez-Cubero et al., 2017).

Multiple studies have reported that NGS technologies are going to be crucial for DNA human typing in cases where forensic specimens and samples are compromised and degraded.

Zhan et al. evaluated the performance of Illumina's MiSeq FGx System in recovering forensic genomic sequencing data from degraded and low-template DNA. The analysis included a sensitivity study with a range of 1 ng−5 pg of 2800 M DNA. By incubating 2800 M DNA at 98°C for 10, 20, 30, 40, 50, 60, 70, 120, and 180 min, DNA was artificially and systematically degraded (resulting in degradation index values ranging from 0.837 to 232.247) (Zhang et al., 2018). When the DNA input was greater than 50 pg or the degradation index was less than 72.28, the detected allele call frequencies were greater than 80%. When the samples were heated for more than 120 minutes or the input quantity was less than 25 pg, the allele balance was less than 0.6. The data also showed that the stutter type and ratio depend on the specific locus, with the sequencing run being the most important factor in noise occurrence. The findings of this study will aid laboratories in developing workflows for the analysis of difficult samples using the ForenSeq kit and the MiSeq FGx system (Zhang et al., 2018).

On heavily degraded DNA samples, NGS appears to offer very promising results. To test the performance and reliability of the ForenSeq DNA Signature Prep kit using the MiSeq sequencer (Illumina), a trial set of DNA samples was artificially degraded by progressive aqueous hydrolysis and analyzed along with the corresponding unmodified DNA sample and control sample (2800 M). In fact, it was able to give important results with three STRs (one autosomal and two Y-chromosome STRs) and 10 autosomal SNPs when the CE was unable to produce a readable profile for the most degraded trial sample. Furthermore, the 1/random match probability value (that is, the likelihood ratio or LR) calculated for the autosomal loci gave a result of 1.7×105, which can be considered an excellent achievement for such a highly damaged DNA sample (Fattorini et al., 2017).

Elwick et al. assessed and sequenced bone and tooth samples exposed to a variety of DNA insults (cremation, embalming, decomposition, thermal degradation, and fire) using Precision ID chemistry and a custom AmpliSeq STR and iiSNP panel on the Ion S5 System, the ForenSeq DNA Signature Prep Kit on the MiSeq FGx system, and the GlobalFiler PCR Amplification Kit on the (Kyleen Elwick et al., 2019). The results showed that traditional CE-based genotyping performed as expected, producing a partial or full DNA profile for all samples, and that both sequencing chemistries and platforms recovered sufficient STR and SNP information from the majority of the same difficult samples.

Despite the degree of damage in some samples, run metrics such as profile completeness and mean read depth produced good results with each system. The majority of sample insults (except decomposed) produced a similar number of alleles for MPS systems (Kyleen Elwick et al., 2019).

Zavala et al. conducted a modeling study that shows that SNPs can increase the significance of identification when analyzing DNA down to an average size of 100 bps for input amounts ranging from 0.375 to 1 ng of nuclear DNA. The findings of this study were then compared to human skeletal material results (n = 14, ninth to eighteenth centuries), demonstrating the effectiveness of the ForenSeq kit for degraded samples. The Promega PowerSeq Mito System was also tested for reliability with human skeletal remains (n = 70, ninth to eighteenth centuries), resulting in successful coverage of 99.29% of the mtDNA control region at 50 coverage or more. This was accompanied by changes to a common DNA extraction technique for skeletal remains that improved template recovery for shorter templates (Zavala et al., 2019).

Developing an NGS assay that can capture mitochondrial probes is an essential component of this technology because it increases the sensitivity and specificity when dealing with a mixture of degraded samples (Calloway et al., 2014). NGS technology has also seen the development of software that can facilitate genome sequencing (Haddrill, 2021). In this way, NGS technology can overcome the limitations of challenging DNA evidence. The Accel NGS 2 S family of DNA library kits and the Accel NGS 1 S Plus DNA Library Kit are among the NGS technologies that can aid the sequencing of damaged and degraded DNA (Regnault et al., 2021).

Elwick compared the baseline tolerance of the two sequencing chemistries and platforms to inhibitors commonly found in human remains recovered from missing person cases (Elwick et al., 2018). Both the HID-Ion AmpliSeq Library Kit and ID Panel and the ForenSeq DNA Signature Prep Kit were susceptible to humic acid, melanin, and collagen; however, the ForenSeq kit handled inhibited melanin and collagen more effectively than the AmpliSeq kit. The ForenSeq kit, on the other hand, was resistant to the effects of hematin and calcium, whereas the AmpliSeq kit was severely inhibited by hematin. When using the ForenSeq kit, STRs and SNPs showed the same trend among inhibitors (Elwick et al., 2018).

The ForenSeq DNA Signature Prep kit was additionally used to investigate the effects of two well-known PCR inhibitors, humic acid and hematin. Humic acid and hematin resulted in decreased read numbers as well as specific negative effects on certain markers. Furthermore, it was demonstrated that a common capillary gel electrophoresis-based STR kit can handle at least 200 times more inhibitors than the ForenSeq DNA Signature Prep kit. This suggests that the PCR components could be improved to ensure analytical success for difficult samples, which is required for MPS to be widely used in forensic STR analysis (Sidstedt et al., 2019).

DNA mixture profile

Initial MPS analyses of STR mixtures revealed that, when appropriately applied and interpreted, NGS could be more powerful, sensitive, informative, and perform better in mixture detection and deconvolution than conventional CE fragment analysis (Holland et al., 2011; Calloway et al., 2014; Van Der Gaag et al., 2016; Wang et al., 2017; Köcher et al., 2018; Novroski et al., 2019; Tao et al., 2019; Ragazzo et al., 2020). The bioinformatic framework My-Forensic-Loci-queries (MyFLq) for the analysis of MPS forensic data has been also presented. MyFLq framework was applied to an Illumina MiSeq dataset of a forensic Illumina amplicon library generated from multilocus STR PCR on both single contributor samples and multiple-person DNA mixtures (Van Neste et al., 2014). Certainly, NGS has the potential to improve mixture deconvolution even further by allowing the detection of specific intrarepeat variations within STR allelic variants of the minor contributor(s) in mixture DNA samples. However, these initial research studies also exposed the limitations of based STR analysis in detecting minor alleles due to the amplification of the major allele prior to library preparation (Gill & Haned, 2013; Tytgat et al., 2020; Van Der Gaag et al., 2016).

This method is particularly useful in studying DNA mixtures for its ability to analyze amplicon pools generated through the use of several primer sets on DNA extracts from multiple specimens (Prosser et al., 2016) and to differentiate the contributors through isometric alleles (with the same length but different sequences) and informative variants (SNPs) within the flanking regions of the repeat motif (Tao et al., 2019; Guo et al., 2017). These findings, while preliminary, may help forensic experts in mixture interpretation, and further investigations on this topic will need to be undertaken (Børsting & Morling, 2015).

The previous study confirmed that NGS methods are powerful to detect sequences from the minor contributor in 1:50 and 1:100 mixtures, which is not possible with CE technology (Ragazzo et al., 2020). In these types of mixtures, it's difficult to separate the reads of the minor contributor alleles from the stutters and noise sequences. The presence of minor contributors was detected in saliva and urine samples using Precision ID GlobalFiler NGS STR Panel v2 in both dilutions, confirming the reliability and robustness of the NGS method for human identification in optimal/suboptimal amounts of DNA (Ragazzo et al., 2020). The "Precision ID GlobalFiler NGS STR Panel v2" kit provides greater identification power because 35 markers are analyzed as opposed to the traditional method's 24 markers from the GlobalFiler Amplification Kit.

For instance, Zeng et al. used the PowerSeq Auto System of 22 STR loci, one Y-STR, and the Amelogenin locus for sequence-based STR analysis on the MiSeq platform (Zeng et al., 2015). The research team compared the results obtained for DNA mixtures using the PowerPlex Fusion System (Promega) on a CE platform. Mixtures at different ratios were amplified using a 0.5 ng input of DNA along with mock forensic

casework samples (blood/saliva and saliva/semen mixtures). Overall, the NGS results indicated that the prototype PowerSeq STR multiplex system could provide results that are concordant and comparable in performance with CE fragment analysis for the interpretation of two-person DNA mixtures. The NGS PowerSeq Auto System, however, enabled the detection of more alleles than CE fragment analysis, with partial profiles of minor contributors reported up to a 19:1 ratio (Zeng et al., 2015).

In 2016, van der Gaag et al. published a paper in which they described the effectiveness of the PowerSeq (Promega) system on the MiSeq sequencer on a series of mixture ratios (1:99, 5:95, 10:90, 20:80, 50:50, 80:20, 90:10, 95:5, and 99:1) using a total DNA input varying from 0.5−6 ng, with a minimum of 60 pg of the minor donor (Van Der Gaag et al., 2016). The sequence reads representing the minor contributors down to 1% of the total DNA profile could only be recovered for a few STR loci. This was explained by the presence of extra PCR amplification or sequence error variants for some of the loci that could complicate the interpretation of STR profiles without prior knowledge of the mixture ratio, and also stutter issues (Van Der Gaag et al., 2016).

A year later, Guo et al. tested the performance of the ForenSeq DNA Signature Prep Kit from the MiSeq FGx Forensic Genomics (Illumina) System in a mixture study. Mixtures at various ratios (1:49, 1:19, 1:9, 1:1, 9:1, 19:1, and 49:1) were prepared at 1 ng of input DNA. ForenSeq Kit showed better performance than conventional CE-STR kits to generate the full profile of the minor contributor, but mixture deconvolution was challenging without using reference DNA profiles (Guo et al., 2017).

In 2018, Köcher et al. published the results of a research project funded by the EU, DNASeqEx (DNA-STR Massive Sequencing & International Information Exchange) investigating the ForenSeq DNA Signature Prep Kit on the MiSeq FGxplatform in a series of male: female ratios of DNA mixtures prepared using 1−2 ng of DNA. The results of this study indicated that the minor DNA profile was detected down to a ratio of 1:20 in all female-male and male-male pairs, while no allele was detected at more extreme ratios (1:1000), using either NGS or CE platforms (Köcher et al., 2018). This study supports evidence from previous observations by Jäger et al. (2017).

Oldoni and Podini (2019) have demonstrated that the sequence-based STR kit (ForenSeq DNA Signature Prep Kit) was more effective in extreme mixtures to report more alleles of the minor donor compared to CE (NGM Select PCR Amplification Kit and PowerPlex Y23 System). Contrary to expectations, this study found that the length-based PowerPlex Y23 kit was more performant to detect most of the male minor alleles at severely underrepresented male mixture contributions compared to the comprehensive MPS multiplex kit (autosomal STRs, Y-STRs, and X-STRs, and individual identification SNPs) incapable to report them. A possible explanation for this might be the potential primer competition (preferential amplification) between female and male donors (Oldoni & Podini, 2019).

Wang et al. evaluated the performance of the Precision ID GlobalFiler NGS 32-plex STR Panel (Thermo Fisher Scientific) on the Ion S5 XL system to investigate a series of mixtures with a total amount of 1 ng. The study confirmed the benefit of mega multiplex MPS kits, which include multiple classes of DNA markers to detect minor alleles compared to conventional CE-STR typing kits (Wang et al., 2017). In this study, the detection of all non-overlapping minor alleles was possible at 1:4 and 1:9 ratios, while partial minor STR profiles (8–10 of the 25 non overlapping alleles) were detected at an approximately 1:19 ratio. However, no minor alleles were detected at most loci in higher-order mixtures (1:29 or 29:1) (Wang et al., 2017).

In addition, NGS can increase the resolution of mtDNA analysis by detecting low-level mixtures such as heteroplasmy ($< 5\%$), which would not be detectable using standard analyses of mtDNA polymorphism by Sanger sequencing with its limit of detection at $\sim$ 10%–20% (Bär et al., 2000; Mamanova et al., 2010; Tsiatis et al., 2010; Parson et al., 2016; Just et al., 2015). Previous research has demonstrated the sensitivity of NGS in detecting and quantifying heteroplasmic variants and mixture components at very low levels (1:250), which may allow for mixture deconvolution (Holland et al., 2011). The NGS study also demonstrated that Y chromosome sequencing could also distinguish between mixed samples of multiple males belonging to the same parent (Yang et al., 2014).

Conclusion: The main challenges of NGS in forensic applications

NGS technologies have contributed to MPS, leading to improved evidence generation for forensic purposes (Ballard et al., 2020). One of the most significant advantages of NGS is that all (or almost all) of the PCR-CE assays may be integrated into a single NGS assay if a capture for the appropriate loci can be developed. In cases where additional studies are required, one NGS assay with several different markers will save time and potentially reduce the time a sample is examined in the laboratory (Børsting & Morling, 2016).

Implementing NGS into forensic workflows is becoming more possible, especially since this technology is getting more affordable and with the development of bioinformatics tools to analyze the huge number of complex data produced (Haddrill, 2021). However, NGS technology has its challenges that extend to forensic science, making it imperative to address them. Molecular aberrations and the overwhelming number of (new) variants that are detected are issues that scientists face when they use these methods to obtain information (Cottrell, 2018), as well as the variable performance of some markers in terms of coverage and locus imbalance (Ballard et al., 2020) and susceptibility to PCR inhibitors when compared with standard STR typing kits (Elwick et al., 2018; Phillips et al., 2018; Sidstedt et al., 2019; Young et al., 2019). Sometimes, the sophisticated bioinformatics systems needed can be time-consuming and costly to obtain.

The goal of the analysis of samples in forensic cases is to determine the identity of the individual(s) who contributed to the sample and possibly estimate any phenotypical characteristics of the individual(s) or identify the specific tissue type(s) in the sample. This requires relatively few markers, and a capture-based approach will be much more economical and require less sample material. However, some NGS systems can only sequence genomes the size of the human genome, and in normal shotgun sequencing operations, substantial regions of the genome are only covered by a low number of reads, which leads to a high risk of misinterpretation of sequencing results, as well as a lack of reproducibility in concordance studies (Børsting & Morling, 2015). Exome sequencing may reveal variants in other genes that may be disease-related. A number of ethical and legislative considerations are raised by introducing NGS in forensic genetics How to handle this information is challenging for clinicians and forensic genetics specialists, too. How much and which loci to sequence, what to report, and whether ignoring sequence information is prudent are certainly critical questions to ask and require the appropriate answers (Børsting & Morling, 2016).

References

Alaeddini, R. (2012). Forensic implications of PCR inhibition - a review. *Forensic Science International: Genetics, 6*(3), 297–305. https://doi.org/10.1016/j.fsigen.2011.08.006

Alonso, A., Barrio, P. A., Müller, P., Köcher, S., Berger, B., Martin, P., Bodner, M., Willuweit, S., Parson, W., Roewer, L., & Budowle, B. (2018). Current state-of-art of STR sequencing in forensic genetics. *Electrophoresis, 39*(21), 2655–2668. https://doi.org/10.1002/elps.201800030

Alvarez-Cubero, M. J., Saiz, M., Martínez-García, B., Sayalero, S. M., Entrala, C., Lorente, J. A., & Martinez-Gonzalez, L. J. (2017). Next generation sequencing: An application in forensic sciences? *Annals of Human Biology, 44*(7), 581–592. https://doi.org/10.1080/03014460.2017.1375155

Álvarez-Iglesias, V., Jaime, J. C., Carracedo, A., & Salas, A. (2007). Coding region mitochondrial DNA SNPs: Targeting east Asian and native American haplogroups. *Forensic Science International: Genetics, 1*(1), 44–55. https://doi.org/10.1016/j.fsigen.2006.09.001

Amankwaa, A. O., & McCartney, C. (2021). The effectiveness of the current use of forensic DNA in criminal investigations in England and Wales. *WIREs Forensic Science, 3*(6). https://doi.org/10.1002/wfs2.1414

Ballantyne, J. (2006). *Assessment and in vitro repair of damaged DNA templates*. DOJ Report.

Ballantyne, K. N., van Oorschot, R. A. H., & Mitchell, R. J. (2007). Comparison of two whole genome amplification methods for STR genotyping of LCN and degraded DNA samples. *Forensic Science International, 166*(1), 35–41. https://doi.org/10.1016/j.forsciint.2006.03.022

Ballard, D., Winkler-Galicki, J., & Wesoły, J. (2020). Massive parallel sequencing in forensics: Advantages, issues, technicalities, and prospects. *International Journal of Legal Medicine, 134*(4), 1291–1303. https://doi.org/10.1007/s00414-020-02294-0

Barba, E. (n.d.). *Chapter 8 - genome informatics pipelines and genome browsers* (pp. 149–169), Academic Press.

Bär, W., Brinkmann, B., Budowle, B., Carracedo, A., Gill, P., Holland, M., Lincoln, P. J., Mayr, W., Morling, N., Olaisen, B., Schneider, P. M., Tully, G., & Wilson, M. (2000). Guidelines for mitochondrial DNA typing. *Vox Sanguinis, 79*(2), 121–125. https://doi.org/10.1159/000031227

Bender, K., Farfán, M. J., & Schneider, P. M. (2004). Preparation of degraded human DNA under controlled conditions. *Forensic Science International, 139*(2–3), 135–140. https://doi.org/10.1016/j.forsciint.2003.10.003

Bieber, F. R., Buckleton, J. S., Budowle, B., Butler, J. M., & Coble, M. D. (2016). Evaluation of forensic DNA mixture evidence: Protocol for evaluation, interpretation, and statistical calculations using the combined probability of inclusion. *BMC Genetics, 17*(1). https://doi.org/10.1186/s12863-016-0429-7

Bogas, V., Carvalho, M., Anjos, M. J., Pinheiro, M. F., & Corte-Real, F. (2009). Genetic identification of degraded DNA samples buried in different types of soil. *Forensic Science International: Genetics Supplement Series, 2*(1), 169–171. https://doi.org/10.1016/j.fsigss.2009.08.031

Børsting, C., & Morling, N. (2015). Next generation sequencing and its applications in forensic genetics. *Forensic Science International: Genetics, 18*, 78–89. https://doi.org/10.1016/j.fsigen.2015.02.002

Børsting, C., & Morling, N. (2016). Genomic applications in forensic medicine. In *Medical and health genomics* (pp. 295–309). Elsevier Inc. https://doi.org/10.1016/B978-0-12-420196-5.00022-8

Borsuk, L. A., Gettings, K. B., Steffen, C. R., Kiesler, K. M., & Vallone, P. M. (2018). Sequence-based US population data for the SE33 locus. *Electrophoresis, 39*(21), 2694–2701. https://doi.org/10.1002/elps.201800091

Braslavsky, I., Hebert, B., Kartalov, E., & Quake, S. R. (2003). Sequence information can be obtained from single DNA molecules. *Proceedings of the National Academy of Sciences, 100*(7), 3960–3964. https://doi.org/10.1073/pnas.0230489100

Brenner, C. H., & Weir, B. S. (2003). Issues and strategies in the DNA identification of world trade center victims. *Theoretical Population Biology, 63*(3), 173–178. https://doi.org/10.1016/S0040-5809(03)00008-X

Bright, J. A., Taylor, D., Gittelson, S., & Buckleton, J. (2017). The paradigm shift in DNA profile interpretation. *Forensic Science International: Genetics, 31*, e24–e32. https://doi.org/10.1016/j.fsigen.2017.08.005

Bruijns, B., Tiggelaar, R., & Gardeniers, H. (2018). Massively parallel sequencing techniques for forensics: A review. *Electrophoresis, 39*(21), 2642–2654. https://doi.org/10.1002/elps.201800082

Budowle, B. (2004). SNP typing strategies. *Forensic Science International, 146*, 139–142.

Budowle, B., Planz, J., Campbell, R., & Eisenberg, A. (2004). Single nucleotide polymorphisms and microarray technology in forensic genetics-development and application to mitochondrial DNA forensic science review. *Forensic Science Review, 16*, 21–36. https://pubmed.ncbi.nlm.nih.gov/26256811/.

Budowle, B., & Van Daal, A. (2008). Forensically relevant SNP classes. *Biotechniques, 44*(5), 603–610. https://doi.org/10.2144/000112806

Budowle, B., Wilson, M. R., DiZinno, J. A., Stauffer, C., Fasano, M. A., Holland, M. M., & Monson, K. L. (1999). Mitochondrial DNA regions HVI and HVII population data. *Forensic Science International, 103*(1), 23–35. https://doi.org/10.1016/S0379-0738(99)00042-0

Burger, J., Hummel, S., Herrmann, B., & Henke, W. (1999). DNA preservation: A microsatellite-DNA study on ancient skeletal remains. *Electrophoresis, 20*(8), 1722–1728. https://doi.org/10.1002/(SICI)1522-2683(19990101)20:8<1722::AID-ELPS1722>3.3.CO;2-W

Butler, J. M. (2012). Advanced topics in forensic DNA typing: Methodology. In *Advanced topics in forensic DNA typing: Methodology*. Elsevier Inc. https://doi.org/10.1016/C2011-0-04189-3

Butler, J. M. (2015). The future of forensic DNA analysis. In *Philosophical transactions of the royal society of London. Series B, biological sciences* (Vol 370).

Butler, J. M., Coble, M. D., & Vallone, P. M. (2007). STRs vs. SNPs: Thoughts on the future of forensic DNA testing. *Forensic Science, Medicine and Pathology, 3*(3), 200–205. https://doi.org/10.1007/s12024-007-0018-1

Butler, J. M., Shen, Y., & McCord, B. R. (2003). The development of reduced size STR amplicons as tools for analysis of degraded DNA. *Journal of Forensic Sciences, 48*(5), 1054–1064. https://doi.org/10.1520/jfs2003043

Butler, J. M., & Willis, S. (2020). Interpol review of forensic biology and forensic DNA typing 2016-2019. *Forensic Science International: Synergy, 2*, 352–367. https://doi.org/10.1016/j.fsisyn.2019.12.002

Calloway, C. D., Kim, H., & Erlich, H. (2014). Resolution of DNA mixtures and analysis of degraded DNA using the 454 DNA sequencing technology. *Report No. 249100*. National Institute of Justice (NIJ). https://www.ojp.gov/ncjrs/virtual-library/abstracts/resolution-dna-mixtures-and-analysis-degraded-dna-using-454-dna.

Cavanaugh, S. E., & Bathrick, A. S. (2018). Direct PCR amplification of forensic touch and other challenging DNA samples: A review. *Forensic Science International: Genetics, 32*, 40–49. https://doi.org/10.1016/j.fsigen.2017.10.005

Chakraborty, R., Stivers, D. N., Su, B., Zhong, Y., & Budowle, B. (1999). The utility of short tandem repeat loci beyond human identification: Implications for development of new DNA typing systems. *Electrophoresis, 20*(8), 1682–1696. https://doi.org/10.1002/(SICI)1522-2683(19990101)20:8<1682::AID-ELPS1682>3.3.CO;2-Q

Chung, D. T., Drábek, J., Opel, K. L., Butler, J. M., & McCord, B. R. (2004). A study on the effects of degradation and template concentration on the amplification efficiency of the STR Miniplex primer sets. *Journal of Forensic Sciences* (Vol 49,(4), 733–740. https://doi.org/10.1520/jfs2003269

Coble, M. D., & Bright, J. A. (2019). Probabilistic genotyping software: An overview. *Forensic Science International: Genetics, 38*, 219–224. https://doi.org/10.1016/j.fsigen.2018.11.009

Colotte, M., Coudy, D., Tuffet, S., & Bonnet, J. (2011). Adverse effect of air exposure on the stability of DNA stored at room temperature. *Biopreservation and Biobanking, 9*(1), 47–50. https://doi.org/10.1089/bio.2010.0028

Constantinescu, C. M., Barbarii, L. E., Iancu, C. B., Constantinescu, A., Neagu, E., Iancu, D., & Girbea, G. (2012). Challenging DNA samples solved with MiniSTR analysis.: Brief overview. *Romanian Journal of Legal Medicine, 20*(1), 51–56. https://doi.org/10.4323/rjlm.2012.51

Cooper, D. N., Smith, B. A., Cooke, H. J., Niemann, S., & Schmidtke, J. (1985). An estimate of unique DNA sequence heterozygosity in the human genome. *Human Genetics, 69*(3), 201–205. https://doi.org/10.1007/BF00293024

Cottrell, P. (2018). *Advantages and drawbacks of next generation sequencing*. SSRN. https://doi.org/10.2139/ssrn.3183340

Dash, H. R., & Das, S. (2018). Microbial degradation of forensic samples of biological origin: Potential threat to human DNA typing. *Molecular Biotechnology, 60*(2), 141–153. https://doi.org/10.1007/s12033-017-0052-5

Diegoli, T. M., Farr, M., Cromartie, C., Coble, M. D., & Bille, T. W. (2012). An optimized protocol for forensic application of the PreCR Repair Mix to multiplex STR amplification of UV-damaged DNA. *Forensic Science International: Genetics, 6*(4), 498–503. https://doi.org/10.1016/j.fsigen.2011.09.003

Divne, A. M., & Allen, M. (2005). A DNA microarray system for forensic SNP analysis. *Forensic Science International, 154*(2–3), 111–121. https://doi.org/10.1016/j.forsciint.2004.09.134

Dixon, L. A., Dobbins, A. E., Pulker, H. K., Butler, J. M., Vallone, P. M., Coble, M. D., Parson, W., Berger, B., Grubwieser, P., Mogensen, H. S., Morling, N., Nielsen, K., Sanchez, J. J., Petkovski, E., Carracedo, A., Sanchez-Diz, P., Ramos-Luis, E., Brion, M., Irwin, J. A., … Gill, P. (2006). Analysis of artificially degraded DNA using STRs and SNPs-results of a collaborative European (EDNAP) exercise. *Forensic Science International, 164*(1), 33–44. https://doi.org/10.1016/j.forsciint.2005.11.011

Elwick, Kyleen, Bus, M. M., King, J. L., Chang, J., Hughes-Stamm, S., & Budowle, B. (2019). Utility of the Ion S5 and MiSeq FGx sequencing platforms to characterize challenging human remains. *Legal Medicine, 41*, 101623. https://doi.org/10.1016/j.legalmed.2019.08.001

Elwick, K., Zeng, X., King, J., Budowle, B., & Hughes-Stamm, S. (2018). Comparative tolerance of two massively parallel sequencing systems to common PCR inhibitors. *International Journal of Legal Medicine, 132*(4), 983–995. https://doi.org/10.1007/s00414-017-1693-4

Evans, J., Maccabee, M., Hatahet, Z., Courcelle, J., Bockrath, R., Ide, H., & Wallace, S. (1993). Thymine ring saturation and fragmentation products: Lesion bypass, misinsertion and implications for mutagenesis. *Mutation Research: Genetic Toxicology, 299*(3–4), 147–156. https://doi.org/10.1016/0165-1218(93)90092-R

Fattorini, P., Previderé, C., Carboni, I., Marrubini, G., Sorçaburu-Cigliero, S., Grignani, P., Bertoglio, B., Vatta, P., & Ricci, U. (2017). Performance of the ForenSeqTM DNA Signature Prep kit on highly degraded samples. *Electrophoresis, 38*(8), 1163–1174. https://doi.org/10.1002/elps.201600290

Fondevila, M., Phillips, C., Naveran, N., Fernandez, L., Cerezo, M., Salas, A., Carracedo, A., & Lareu, M. V. (2008). Case report: Identification of skeletal remains using short-amplicon marker analysis of severely degraded DNA extracted from a decomposed and charred femur. *Forensic Science International: Genetics, 2*(3), 212–218. https://doi.org/10.1016/j.fsigen.2008.02.005

Gill, P. (2001). An assessment of the utility of single nucleotide polymorphisms (SNPs) for forensic purposes. *International Journal of Legal Medicine, 114*(4–5), 204–210. https://doi.org/10.1007/s004149900117

Gill, P., & Haned, H. (2013). A new methodological framework to interpret complex DNA profiles using likelihood ratios. *Forensic Science International: Genetics,* 7(2), 251–263. https://doi.org/10.1016/j.fsigen.2012.11.002

Gill, P., Haned, H., Bleka, O., Hansson, O., Dørum, G., & Egeland, T. (2015). Genotyping and interpretation of STR-DNA: Low-template, mixtures and database matches - twenty years of research and development. *Forensic Science International: Genetics, 18*, 100–117. https://doi.org/10.1016/j.fsigen.2015.03.014

Gill, P., Kirkham, A., & Curran, J. (2007). LoComatioN: A software tool for the analysis of low copy number DNA profiles. *Forensic Science International, 166*(2–3), 128–138. https://doi.org/10.1016/j.forsciint.2006.04.016

Gill, P., Werrett, D. J., Budowle, B., & Guerrieri, R. (2004). An assessment of whether SNPs will replace STRs in national DNA databases - joint considerations of the DNA working group of the European network of forensic science institutes (ENFSI) and the scientific working group on DNA analysis methods (SWGDAM). *Science and Justice - Journal of the Forensic Science Society, 44*(1), 51–53. https://doi.org/10.1016/S1355-0306(04)71685-8

Goodsell, D. S. (2001). The molecular perspective: Ultraviolet light and pyrimidine dimers. *Stem Cells, 19*(4), 348–349. https://doi.org/10.1634/stemcells.19-4-348

Gršković, B., Zrnec, D., Popović, M., Petek, M. J., Primorac, D., & Mršić, G. (2013). Effect of ultraviolet c radiation on biological samples. *Croatian Medical Journal, 54*(3), 263–271. https://doi.org/10.3325/cmj.2013.54.263

Guo, F., Yu, J., Zhang, L., & Li, J. (2017). Massively parallel sequencing of forensic STRs and SNPs using the Illumina ForenSeq DNA signature Prep kit on the MiSeq FGx^{TM} forensic genomics system. *Forensic Science International: Genetics, 31*, 135–148. https://doi.org/10.1016/j.fsigen.2017.09.003

Haddrill, P. R. (2021). Developments in forensic dna analysis. *Emerging Topics in Life Sciences, 5*(3), 381–393. https://doi.org/10.1042/ETLS20200304

Hall, A., & Ballantyne, J. (2004). Characterization of UVC-induced DNA damage in bloodstains: Forensic implications. *Analytical and Bioanalytical Chemistry, 380*(1), 72–83. https://doi.org/10.1007/s00216-004-2681-3

Harris, T. D., Buzby, P. R., Babcock, H., Beer, E., Bowers, J., Braslavsky, I., Causey, M., Colonell, J., DiMeo, J., Efcavitch, J. W., Giladi, E., Gill, J., Healy, J., Jarosz, M., Lapen, D., Moulton, K., Quake, S. R., Steinmann, K., Thayer, E., … Xie, Z. (2008). Single-molecule DNA sequencing of a viral genome. *Science, 320*(5872), 106–109. https://doi.org/10.1126/science.1150427

Hedman, J., & Rådström, P. (2013). Overcoming inhibition in real-time diagnostic PCR. *Methods in Molecular Biology, 943*, 17–48. https://doi.org/10.1007/978-1-60327-353-4_2

Holden, A. L. (2002). The SNP Consortium: Summary of a private consortium effort to develop an applied map of the human genome. *Biotechniques, 32*(6), S22–S26. https://doi.org/10.2144/jun0203

Holland, M. M., Cave, C. A., Holland, C. A., & Bille, T. W. (2003). Development of a quality, high throughput DNA analysis procedure for skeletal samples to assist with the identification of victims from the World Trade Center attacks. *Croatian Medical Journal, 44*(3), 264–272.

Holland, M. M., McQuillan, M. R., & O'Hanlon, K. A. (2011). Second generation sequencing allows for mtDNA mixture deconvolution and high resolution detection of heteroplasmy. *Croatian Medical Journal, 52*(3), 299–313. https://doi.org/10.3325/cmj.2011.52.299

Holt, C., Walichiewicz, P., Eagles, J., Daulo, A., Didier, M., Edwards, C., Fleming, K., Han, Y., Hill, T., Li, S., Rensfield, A., Sa, D., & Stephens, K. (2019). Mitochondrial DNA data analysis strategies that inform MPS-based forensic casework implementation. *Forensic Science International: Genetics Supplement Series,* 7(1), 389–391. https://doi.org/10.1016/j.fsigss.2019.10.024

Jäger, A. C., Alvarez, M. L., Davis, C. P., Guzmán, E., Han, Y., Way, L., Walichiewicz, P., Silva, D., Pham, N., Caves, G., Bruand, J., Schlesinger, F., Pond, S. J. K., Varlaro, J., Stephens, K. M., & Holt, C. L. (2017). Developmental validation of the MiSeq FGx forensic genomics system for targeted next generation sequencing in forensic DNA casework and database laboratories. *Forensic Science International: Genetics, 28*, 52–70. https://doi.org/10.1016/j.fsigen.2017.01.011

Jordan, D., & Mills, D. (2021). Past, present, and future of DNA typing for analyzing human and non-human forensic samples. *Frontiers in Ecology and Evolution, 9*, 646130.

Just, R. S., Irwin, J. A., & Parson, W. (2015). Mitochondrial DNA heteroplasmy in the emerging field of massively parallel sequencing. *Forensic Science International: Genetics, 18*, 131–139. https://doi.org/10.1016/j.fsigen.2015.05.003

Kayser, M., & Parson, W. (2018). Transitioning from forensic genetics to forensic genomics. *Genes, 9*(1). https://doi.org/10.3390/genes9010003

Kelly, H., Bright, J. A., Buckleton, J. S., & Curran, J. M. (2014). A comparison of statistical models for the analysis of complex forensic DNA profiles. *Science & Justice, 54*(1), 66–70. https://doi.org/10.1016/j.scijus.2013.07.003

Kelly, H., Bright, J. A., Curran, J., & Buckleton, J. (2012). The interpretation of low level DNA mixtures. *Forensic Science International: Genetics, 6*(2), 191–197. https://doi.org/10.1016/j.fsigen.2011.04.013

Klein, R., Neumann, C., & Roy, R. (2015). Detection of insertion/deletion polymorphisms from challenged samples using the Investigator DIPplex® Kit. *Forensic Science International: Genetics, 16*, 29–37. https://doi.org/10.1016/j.fsigen.2014.11.020

de Knijff, P. (2019). From next generation sequencing to now generation sequencing in forensics. *Forensic Science International: Genetics, 38*, 175–180. https://doi.org/10.1016/j.fsigen.2018.10.017

Köcher, S., Müller, P., Berger, B., Bodner, M., Parson, W., Roewer, L., & Willuweit, S. (2018). Interlaboratory validation study of the ForenSeq™ DNA signature Prep kit. *Forensic Science International: Genetics, 36*, 77–85. https://doi.org/10.1016/j.fsigen.2018.05.007

Lindahl, T. (1993). Instability and decay of the primary structure of DNA. *Nature, 362*(6422), 709–715. https://doi.org/10.1038/362709a0

Llamas, B., Valverde, G., Fehren-Schmitz, L., Weyrich, L. S., Cooper, A., & Haak, W. (2017). From the field to the laboratory: Controlling DNA contamination in human ancient DNA research in the high-throughput sequencing era. *Science and Technology of Archaeological Research, 3*(1), 1–14. https://doi.org/10.1080/20548923.2016.1258824

Machado, H., & Granja, R. (2020). *DNA technologies in criminal investigation and courts* (pp. 45–56). Springer Science and Business Media LLC. https://doi.org/10.1007/978-981-15-2429-5_4

Mamanova, L., Coffey, A. J., Scott, C. E., Kozarewa, I., Turner, E. H., Kumar, A., Howard, E., Shendure, J., & Turner, D. J. (2010). Target-enrichment strategies for next-generation sequencing. *Nature Methods, 7*(2), 111–118. https://doi.org/10.1038/nmeth.1419

Mateen, R. M., Tariq, A., & Hussain, M. (2020). Generating DNA profile from low copy number DNA: Strategies and associated risks. *Acta Scientiarum Biological Sciences, 42*, 1–7. https://doi.org/10.4025/actascibiolsci.v42i1.52239

Mautner, M. E., Pérez Santángelo, A., Corti Bielsa, R. M., Sala, A., Ginart, S., Corach, D., & Kalendar, R. (2017). Using long ssDNA polynucleotides to amplify STRs loci in degraded DNA samples. *PLoS One, 12*(11), e0187190. https://doi.org/10.1371/journal.pone.0187190

McCord, B. R., Gauthier, Q., Cho, S., Roig, M. N., Gibson-Daw, G. C., Young, B., Taglia, F., Zapico, S. C., Mariot, R. F., Lee, S. B., & Duncan, G. (2019). Forensic DNA analysis. *Analytical Chemistry, 91*(1), 673–688. https://doi.org/10.1021/acs.analchem.8b05318

Mechta, M., Ingerslev, L. R., Fabre, O., Picard, M., & Barrès, R. (2017). Evidence suggesting absence of mitochondrial DNA methylation. *Frontiers in Genetics, 8*. https://doi.org/10.3389/fgene.2017.00166

Meiklejohn, K. A., & Robertson, J. M. (2017). Evaluation of the precision ID identity panel for the ion Torrent™ PGM™ sequencer. *Forensic Science International: Genetics, 31*, 48–56. https://doi.org/10.1016/j.fsigen.2017.08.009

Mishmar, D., Levin, R., Naeem, M. M., & Sondheimer, N. (2019). Higher order organization of the mtDNA: Beyond mitochondrial transcription factor A. *Frontiers in Genetics, 10*. https://doi.org/10.3389/fgene.2019.01285

Montano, E. A., Bush, J. M., Garver, A. M., Larijani, M. M., Wiechman, S. M., Baker, C. H., Wilson, M. R., Guerrieri, R. A., Benzinger, E. A., Gehres, D. N., & Dickens, M. L. (2018). Optimization of the Promega PowerSeq™ Auto/Y system for efficient integration within a forensic DNA laboratory. *Forensic Science International: Genetics, 32*, 26–32. https://doi.org/10.1016/j.fsigen.2017.10.002

Moreno, L. I., Galusha, M. B., & Just, R. (2018). A closer look at Verogen's ForenSeq™ DNA Signature Prep kit autosomal and Y-STR data for streamlined analysis of routine reference samples. *Electrophoresis, 39*(21), 2685–2693. https://doi.org/10.1002/elps.201800087

Novroski, N. M. M., Wendt, F. R., Woerner, A. E., Bus, M. M., Coble, M., & Budowle, B. (2019). Expanding beyond the current core STR loci: An exploration of 73 STR markers with increased diversity for enhanced DNA mixture deconvolution. *Forensic Science International: Genetics, 38*, 121–129. https://doi.org/10.1016/j.fsigen.2018.10.013

Oldoni, F., & Podini, D. (2019). Forensic molecular biomarkers for mixture analysis. *Forensic Science International: Genetics, 41*, 107–119. https://doi.org/10.1016/j.fsigen.2019.04.003

Opel, K. L., Chung, D. T., Drábek, J., Butler, J. M., & McCord, B. R. (2007). Developmental validation of reduced-size STR miniplex primer sets. *Journal of Forensic Sciences, 52*(6), 1263–1271. https://doi.org/10.1111/j.1556-4029.2007.00584.x

Parson, W., Ballard, D., Budowle, B., Butler, J. M., Gettings, K. B., Gill, P., Gusmão, L., Hares, D. R., Irwin, J. A., King, J. L., Knijff, P. D., Morling, N., Prinz, M., Schneider, P. M., Neste, C. V., Willuweit, S., & Phillips, C. (2016). Massively parallel sequencing of forensic STRs: Considerations of the DNA commission of the International Society for Forensic Genetics (ISFG) on minimal nomenclature requirements. *Forensic Science International: Genetics, 22*, 54–63. https://doi.org/10.1016/j.fsigen.2016.01.009

Parson, W., Gusmão, L., Hares, D. R., Irwin, J. A., Mayr, W. R., Morling, N., Pokorak, E., Prinz, M., Salas, A., Schneider, P. M., & Parsons, T. J. (2014). DNA commission of the international society for forensic genetics: Revised and extended guidelines for mitochondrial DNA typing. *Forensic Science International: Genetics, 13*, 134–142. https://doi.org/10.1016/j.fsigen.2014.07.010

Peck, M. A., Sturk-Andreaggi, K., Thomas, J. T., Oliver, R. S., Barritt-Ross, S., & Marshall, C. (2018). Developmental validation of a Nextera XT mitogenome Illumina MiSeq sequencing method for high-quality samples. *Forensic Science International: Genetics, 34*, 25–36. https://doi.org/10.1016/j.fsigen.2018.01.004

Pereira, V., Longobardi, A., & Børsting, C. (2018). Sequencing of mitochondrial genomes using the precision ID mtDNA whole genome panel. *Electrophoresis, 39*(21), 2766–2775. https://doi.org/10.1002/elps.201800088

Pereira, R., Phillips, C., Alves, C., Amorim, A., Carracedo, A., & Gusmao, L. (2009). A new multiplex for human identification using insertion/deletion polymorphisms. *Electrophoresis, 30*(21), 3682–3690. https://doi.org/10.1002/elps.200900274

Phillips, C., Devesse, L., Ballard, D., van Weert, L., de la Puente, M., Melis, S., Álvarez Iglesias, V., Freire-Aradas, A., Oldroyd, N., Holt, C., Syndercombe Court, D., Carracedo, Á., & Lareu, M. V. (2018). Global patterns of STR sequence variation: Sequencing the CEPH human genome diversity panel for 58 forensic STRs using the Illumina ForenSeq DNA Signature Prep Kit. *Electrophoresis, 39*(21), 2708–2724. https://doi.org/10.1002/elps.201800117

Phillips, C., Gettings, K. B., King, J. L., Ballard, D., Bodner, M., Borsuk, L., & Parson, W. (2018). The devil's in the detail": Release of an expanded, enhanced and dynamically revised forensic STR Sequence Guide. *Forensic Science International: Genetics, 34*, 162–169. https://doi.org/10.1016/j.fsigen.2018.02.017

Prosser, S. W. J., Dewaard, J. R., Miller, S. E., & Hebert, P. D. N. (2016). DNA barcodes from century-old type specimens using next-generation sequencing. *Molecular Ecology Resources, 16*(2), 487–497. https://doi.org/10.1111/1755-0998.12474

Ragazzo, M., Carboni, S., Caputo, V., Buttini, C., Manzo, L., Errichiello, V., Puleri, G., & Giardina, E. (2020). Interpreting mixture profiles: Comparison between precision ID GlobalFiler™ NGS STR panel v2 and traditional methods. *Genes, 11*(6), 591. https://doi.org/10.3390/genes11060591

Regnault, B., Bigot, T., Ma, L., Pérot, P., Temmam, S., & Eloit, M. (2021). Deep impact of random amplification and library construction methods on viral metagenomics results. *Viruses, 13*(2), 253. https://doi.org/10.3390/v13020253

Sachidanandam, R., Weissman, D., Schmidt, S. C., Kakol, J. M., Stein, L. D., Marth, G., Sherry, S., Mullikin, J. C., Mortimore, B. J., Willey, D. L., Hunt, S. E., Cole, C. G., Coggill, P. C., Rice, C. M., Ning, Z., Rogers, J., Bentley, D. R., Kwok, P. Y., Mardis, E. R., … Altshuler, D.

(2001). A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature, 409*(6822), 928–933. https://doi.org/10.1038/35057149

Schrader, C., Schielke, A., Ellerbroek, L., & Johne, R. (2012). PCR inhibitors - occurrence, properties and removal. *Journal of Applied Microbiology, 113*(5), 1014–1026. https://doi.org/10.1111/j.1365-2672.2012.05384.x

Shih, S., Bose, N., Gonçalves, A., Erlich, H., & Calloway, C. (2018). Applications of probe capture enrichment next generation sequencing for whole mitochondrial genome and 426 nuclear SNPs for forensically challenging samples. *Genes, 9*(1), 49. https://doi.org/10.3390/genes9010049

Sidstedt, M., Steffen, C. R., Kiesler, K. M., Vallone, P. M., Rådström, P., & Hedman, J. (2019). The impact of common PCR inhibitors on forensic MPS analysis. *Forensic Science International: Genetics, 40*, 182–191. https://doi.org/10.1016/j.fsigen.2019.03.001

Slatkin, M., & Racimo, F. (2016). Ancient DNA and human history. *Proceedings of the National Academy of Sciences, 113*(23), 6380–6387. https://doi.org/10.1073/pnas.1524306113

Sobrino, B., Brión, M., & Carracedo, A. (2005). SNPs in forensic genetics: A review on SNP typing methodologies. *Forensic Science International, 154*(2–3), 181–194. https://doi.org/10.1016/j.forsciint.2004.10.020

Tan, E. (2010). *Sample collection system for DNA analysis of forensic evidence: Towards practical, fully-integrated STR analysis* (pp. 1–45). NetBio Inc.

Tao, R., Qi, W., Chen, C., Zhang, J., Yang, Z., Song, W., Zhang, S., & Li, C. (2019). Pilot study for forensic evaluations of the Precision ID GlobalFiler™ NGS STR Panel v2 with the Ion S5™ system. *Forensic Science International: Genetics, 43*, 102147. https://doi.org/10.1016/j.fsigen.2019.102147

Templeton, J. E. L., Brotherton, P. M., Llamas, B., Soubrier, J., Haak, W., Cooper, A., & Austin, J. J. (2013). DNA capture and next-generation sequencing can recover whole mitochondrial genomes from highly degraded samples for human identification. *Investigative Genetics, 4*(1). https://doi.org/10.1186/2041-2223-4-26

Travers, A., & Muskhelishvili, G. (2015). DNA structure and function. *FEBS Journal, 282*(12), 2279–2295. https://doi.org/10.1111/febs.13307

Tsiatis, A. C., Norris-Kirby, A., Rich, R. G., Hafez, M. J., Gocke, C. D., Eshleman, J. R., & Murphy, K. M. (2010). Comparison of Sanger sequencing, pyrosequencing, and melting curve analysis for the detection of KRAS mutations: Diagnostic and clinical implications. *Journal of Molecular Diagnostics, 12*(4), 425–432. https://doi.org/10.2353/jmoldx.2010.090188

Tytgat, O., Gansemans, Y., Weymaere, J., Rubben, K., Deforce, D., & Van Nieuwerburgh, F. (2020). Nanopore sequencing of a forensic STR multiplex reveals loci suitable for single-contributor STR profiling. *Genes, 11*(4), 381. https://doi.org/10.3390/genes11040381

Van Der Gaag, K. J., De Leeuw, R. H., Hoogenboom, J., Patel, J., Storts, D. R., Laros, J. F. J., & De Knijff, P. (2016). Massively parallel sequencing of short tandem repeats - population data and mixture analysis results for the PowerSeq™ system. *Forensic Science International: Genetics, 24*, 86–96. https://doi.org/10.1016/j.fsigen.2016.05.016

Van Neste, C., Vandewoestyne, M., Van Criekinge, W., Deforce, D., & Van Nieuwerburgh, F. (2014). MyForensic-Loci-queries (MyFLq) framework for analysis of forensic STR data generated by massive parallel sequencing. *Forensic Science International: Genetics, 9*(1), 1–8. https://doi.org/10.1016/j.fsigen.2013.10.012

Van der Schans, G. P. (1978). Gamma-ray induced double-Strand breaks in DNA resulting from randomly-inflicted single-strand breaks: Temporal local denaturation, a new radiation phenomenon? *International Journal of Radiation Biology, 33*(2), 105–120. https://doi.org/10.1080/09553007814550011

Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., Holt, R. A., Gocayne, J. D., Amanatides, P., Ballew, R. M., Huson, D. H., Wortman, J. R., Zhang, Q., Kodira, C. D., Zheng, X. H., Chen, L., … Zhu, X. (2001). The sequence of the human genome. *Science, 291*(5507), 1304–1351. https://doi.org/10.1126/science.1058040

Wang, D. G., Fan, J. B., Siao, C. J., Berno, A., Young, P., Sapolsky, R., Ghandour, G., Perkins, N., Winchester, E., Spencer, J., Kruglyak, L., Stein, L., Hsie, L., Topaloglou, T., Hubbell, E., Robinson, E., Mittmann, M., Morris, M. S., Shen, N., … Lander, E. S. (1998). Large-scale

identification, mapping, and genotyping of single- nucleotide polymorphisms in the human genome. *Science, 280*(5366), 1077–1082. https://doi.org/10.1126/science.280.5366.1077

Wang, Z., Zhou, D., Wang, H., Jia, Z., Liu, J., Qian, X., Li, C., & Hou, Y. (2017). Massively parallel sequencing of 32 forensic markers using the precision ID GlobalFiler™ NGS STR panel and the ion PGM™ system. *Forensic Science International: Genetics, 31*, 126–134. https://doi.org/10.1016/j.fsigen.2017.09.004

Whitaker, J. P., Cotton, E. A., & Gill, P. (2001). A comparison of the characteristics of profiles produced with the AMPFlSTR® SGM Plus™ multiplex system for both standard and low copy number (LCN) STR DNA analysis. *Forensic Science International, 123*(2–3), 215–223. https://doi.org/10.1016/S0379-0738(01)00557-6

Yang, Y., Xie, B., & Yan, J. (2014). Application of next-generation sequencing technology in forensic science. *Genomics, Proteomics & Bioinformatics, 12*(5), 190–197. https://doi.org/10.1016/j.gpb.2014.09.001

Young, B., Faris, T., & Armogida, L. (2019). A nomenclature for sequence-based forensic DNA analysis. *Forensic Science International: Genetics, 42*, 14–20. https://doi.org/10.1016/j.fsigen.2019.06.001

Zavala, E. I., Rajagopal, S., Perry, G. H., Kruzic, I., Bašić, Ž., Parsons, T. J., & Holland, M. M. (2019). Impact of DNA degradation on massively parallel sequencing-based autosomal STR, iiSNP, and mitochondrial DNA typing systems. *International Journal of Legal Medicine, 133*(5), 1369–1380. https://doi.org/10.1007/s00414-019-02110-4

Zeng, X., King, J., Hermanson, S., Patel, J., Storts, D. R., & Budowle, B. (2015). An evaluation of the PowerSeq™ Auto system: A multiplex short tandem repeat marker kit compatible with massively parallel sequencing. *Forensic Science International: Genetics, 19*, 172–179. https://doi.org/10.1016/j.fsigen.2015.07.015

Zhang, Q., Zhou, Z., Liu, Q., Liu, L., Shao, L., Zhang, M., Ding, X., Gao, Y., & Wang, S. (2018). Evaluation of the performance of Illumina's ForenSeq™ system on serially degraded samples. *Electrophoresis, 39*(21), 2674–2684. https://doi.org/10.1002/elps.201800101

Zidkova, A., Horinek, A., Kebrdlova, V., & Korabecna, M. (2013). Application of the new insertion-deletion polymorphism kit for forensic identification and parentage testing on the Czech population. *International Journal of Legal Medicine, 127*(1), 7–10. https://doi.org/10.1007/s00414-011-0649-3

Qiagen. (2022). *Next-Generation Sequencing for HID*. Retrieved January 10, 2023, 2023, from https://www.qiagen.com/us/product-categories/human-id-and-forensics/nextgeneration-sequencing.

Thermo Fisher Scientific. (2022). *Precision ID NGS System for Human Identification: Targeted Sequencing for Forensic DNA Analysis*. https://www.thermofisher.com/sa/en/home/industrial/forensics/human-identification/forensic-dna-analysis/dna-analysis/next-generation-sequencing-ngs-forensics.html.

This page intentionally left blank

CHAPTER 21

Detection of human body fluid through mRNA analysis using NGS

Quentin Gauthier
Alexandria, VA, United States

Introduction

The use of RNA for forensics has seen significant interest in the past two decades (Børsting & Morling, 2015). Although a relative newcomer to forensic genetics, RNA analysis has the potential for a wide variety of use cases not offered by DNA, such as postmortem interval, time since deposition, chronological age, and body fluid identification (Bauer et al., 2003; Børsting & Morling, 2015; Peters et al., 2015; Weinbrecht et al., 2017). Body fluid identification in particular has captured the attention of forensic scientists for its ability to contextualize the presence of biological material at crime scenes and in evidence. The ability to distinguish saliva from semen in a stain can substantiate allegations of sexual assault in a way that the presence of DNA simply cannot. Current methodologies employed by forensic laboratories most commonly include enzymatic assays and microscopic analysis (Scientific Working Group on DNA Analysis Methods (SWGDAM), 2015). Although these methods are relatively cheap and easy to employ, there are a number of downsides associated with them. Some tests, such as the SALIgAE reagent, require 10 μL of saliva for visualization and therefore aren't suitable for trace analysis (Old et al., 2009). The RSID-Semen lateral flow assay is highly specific for semenogelin; however, it suffers from false negatives in the presence of other body fluids commonly found in mixtures in forensic contexts (Old et al., 2012). Still other tests, like the ABAcard HemaTrace are destructive to the portion of the sample tested, which isn't the desired result for any forensic DNA analysis method (Kulstein & Wiegand, 2018).

Other approaches, including DNA methylation, real-time polymerase chain reaction (RT-PCR), and high-resolution melt analysis targeting microRNA have seen significant attention in the past decade (Antunes et al., 2016; Gauthier et al., 2019; Haas et al., 2009; Madi et al., 2012; Park et al., 2014; van den Berge et al., 2014; Zubakov et al., 2010). However, next-generation sequencing for messenger RNA (mRNA) has only recently begun to receive the same attention. Having mRNA as a target for body fluid identification is a logical conclusion, as it has been reported to be a stable molecule that can be easily and reliably recovered from evidence submitted for analysis (Setzer et al., 2008). With approximately 22,000 coding genes in the human genome, not all of the genome is

Next Generation Sequencing (NGS) Technology in DNA Analysis
ISBN 978-0-323-99144-5, https://doi.org/10.1016/B978-0-323-99144-5.00024-X
© 2024 Elsevier Inc.
All rights reserved.

expressed in every cell type (Salzberg, 2018). The presence of mRNA, therefore, can be an indicator of cellular function and an indicator of the bodily origin of the mRNA (Melé et al., 2015). Roughly 12% of RNA transcripts are enriched in a single tissue, 10% are expressed in a subset of tissues at a higher rate than other tissues, and approximately 40% are housekeeping genes that would be expressed in all cells (Karlsson et al., 2021; Uhlén et al., 2015). Another benefit to the analysis of RNA is the ability to coextract DNA and RNA simultaneously (Bowden et al., 2011). With this approach, it is possible to generate a body fluid identification from RNA and a full DNA profile from the same sample of evidence, limiting the possibility of biased sampling (Alvarez et al., 2004). RNA and DNA coextraction utilize both organic- and column-based methods, with both being hands-on approaches rather than automated (Parker et al., 2011; Watanabe & Akutsu, 2020). To analyze extracted mRNA, next-generation sequencing could provide a higher level of specificity, sensitivity, and coverage than many of the previous technologies, such as RT-PCR followed by capillary electrophoresis or rt-quantitative polymerase chain reaction (Bustin, 2000; Juusola & Ballantyne, 2007; Lindenbergh et al., 2012). Efforts have even been made to develop assays to simultaneously analyze DNA and RNA at the same time, a feat made possible by the use of next-generation sequencing (Zubakov et al., 2015). While the number of studies targeting mRNA for body fluid identification using next-generation sequencing remains limited, the results have shown great promise for the field. Recently made efforts have focused on expanding the number of body fluid-specific mRNA transcripts that are known, creating assays for targeted sequencing, and exploring newer sequencing technologies.

Body fluid identification

A multitude of transcripts have been identified and characterized for their ability to identify a variety of forensically relevant body fluids. Typically, saliva, semen, blood, and vaginal epithelia are the primary body fluids encountered; however, mRNA sequencing has found a number of markers for other body fluids, such as sweat, nasal mucosa, menstrual blood, and others (Sakurada et al., 2011). Body fluid identification via transcripts relies on the relative stability and abundance of certain transcripts in tissues, while low abundance and stability transcripts can only be used on very fresh samples (Lin et al., 2015). The use of next-generation sequencing platforms for mRNA sequencing, while relatively new, has already produced a number of promising studies for the identification of body fluids. Table 21.1 lists the markers that have been published in recent next-generation sequencing studies.

In 2015, Zubakov et al. produced the first proof-of-concept study for the targeted sequencing of DNA and RNA simultaneously (Zubakov et al., 2015). In their work, the assay targeted 9 autosomal STRs, amelogenin, and 12 mRNAs for body fluid identification. The mRNA targets were capable of positively identifying blood, menstrual

Table 21.1 Published mRNA body fluid identification markers derived from NGS.

Group	Sequencing platform	Saliva	Blood	Vaginal fluid	Semen	Menstrual blood	Skin/Sweat	Nasal mucosa	Spermatozoa	Housekeeping
Zubakov et al. (2015)	Ion Torrent	HTN3 STATH	ALAS2 SPTB	CYP2B7P1 MUC4	PRM1 TGM4	MMP10 MMP11	CCL27 LCE1C	-	-	HPRT1 SDHA
Lin et al. (2015)	HiSeq2500	STATH HTN3	GYPA	-	-	MMP11	-	-	-	UBE2D2 TCEA1 G6PD
Hanson et al. (2018)	MiSeq FGx	HTN3 HTN1 STATH PRB3 PRB4 PRH2	ALAS2 ANK1 SPTB CD3G CD93 AMICA1	CYP287P1 DKK4 FAM83D CYP2A6	PRM1 PRM2 TGM4 SEMG1 SEMG2 KLK3	MMP10 LEFTY2 MMP7 MMP11 SFRP4	LCE1C CCL27 IL37 SERPINA12 KRT77 COL17A1	-	-	-
Ingold et al. (2018)	Ion Torrent	HTN3 STATH PRB3 PRH2 PRB1 MUC7	ALAS2 ANK1 SPTB CD3G	CYP287P1 DKK4 FAM83D CYP2A6	PRM1 PRM2 TGM4 SEMG1 SEMG2 KLK3	MMP10 LEFTY2 MMP7 SFRP4	LCE1C CCL27 IL37 SERPINA12 KRT77	-	-	-
Ingold et al. (2020)	Ion Torrent	HTN3 HTN1 STATH PRB3 PRB4 PRH2 MUC7	ALAS2 ANK1 SPTB CD3G CD93 AMICA1	CYP287P1 DKK4 FAM83D CYP2A6 CYP2A7	PRM1 PRM2 TGM4 SEMG1 SEMG2 KLK3	MMP10 LEFTY2 MMP7 MMP11 SFRP4	LCE1C CCL27 IL37 SERPINA12 KRT77 COL17A1	-	-	-
Chirnside et al. (2020)	MiSeq FGx	-	-	-	-	-	-	OPRPN BPIFA1	-	-
Albani & Fleming (2018)	HiSeq2500	-	HBD SLC4A1	-	KLK2	MMP3 STC1	-	-	TNP1	-

blood, saliva, semen, vaginal secretions, and skin in single-source samples. Additionally, in 2015, Lin et al. analyzed degraded RNA samples for their ability to be successfully sequenced in NGS applications while providing sufficient coverage for positive identification of body fluids (Lin et al., 2015). In their study, they developed markers targeting saliva, blood, menstrual blood, and vaginal fluid. Each body fluid was successfully identified even after degradation by aging for 6 weeks, resulting in RNA Integrity Numbers falling below eight, the minimum recommended RNA quality for sequencing. In a follow-up study in 2016, Lin et al. described the concept of transcript stable regions (StaRs) (Lin et al., 2016). These regions of the transcriptome were identified by comparing the transcriptomic sequencing coverage of fresh and degraded samples. Lin et al. proposed that future assays for body fluid identification target these StaRs to achieve consistent results regardless of the sample quality. This would be achieved by designing primers highly specific to a small number of body fluid-specific transcriptomic regions. In 2018, Hanson et al. published a targeted mRNA multiplex sequencing approach that targets 33 different transcripts for the identification of saliva, blood, vaginal fluid, semen, menstrual blood, and skin (Hanson et al., 2018). The sensitivity of this assay was shown to be as low as 5−10 ng among the body fluids tested, despite the 50 ng recommended input from the manufacturer of the library preparation kit. Additionally, blinded single source and mixture samples were sent to another laboratory for analysis. Of the 16 samples, 15 were accurately identified, with the 16th sample having the major component accurately identified. Additionally, in 2018, Ingold et al. presented the results of a collaborative experiment using the Hanson 33-plex on the MiSeq FGx platform and an in-house 29-plex on the Ion Torrent S5 system (Ingold et al., 2018). In this study, a dozen laboratories were invited to test both platforms on a standardized set of samples. Although some interlaboratory variability on the read counts was observed, the conclusions of each laboratory were quite similar. In this trial, the S5 workflow demonstrated lower reliability with the low input samples provided to the various laboratories. In 2019, Hanson et al. developed an 11-target assay for mRNA body fluid identification of saliva, blood, and semen containing 21 coding region single nucleotide polymorphisms (cSNPs) (Hanson et al., 2019). Their assay was capable of successfully identifying the three body fluids targeted and differentiating saliva, blood, and semen transcripts from different donors. A follow-up study in 2020 by Ingold et al. once again invited laboratories to compare the Hanson 33-plex assay to an improved 35-plex assay from Ingold et al. that used the previous 29 targets as well as 6 new targets and incorporated cSNPs that are located inside of mRNA transcripts targeting various body fluids (Ingold et al., 2020). The benefit of the 35-plex is that the cSNPs are capable of linking specific individuals to a particular body fluid. However, the performance of the cSNPs in this study was unreliable, leading to ambiguous results overall. The authors recommended pursuing additional, more stable cSNPs in an even larger assay to improve the ability to identify body fluids and associate individuals with a specific body fluid. The overall performance

of the two assays on two different platforms remained comparable and represented the largest targeted multiplex assays currently published.

Other groups have contributed to the region of body fluid identification via mRNA sequencing by identifying markers for other body fluids that, while less common in forensics, are important for understanding the full spectrum of body fluids that could be encountered at a crime scene. Chirnside et al. set out to develop markers for the specific identification of nasal mucosa as a body fluid distinct from saliva and tongue samples, as well as blood, semen, and vaginal fluids (Chirnside et al., 2020). Their initial results showed six possible candidates for nasal mucosa identification, but after further specificity testing against all body fluids, OPRPN and BPIFA1 were determined to be the only two markers that they could identify with sufficient specificity. The justification for identifying nasal mucosa as a body fluid is to eliminate the reported cross-reactivity of mRNA assays for saliva and nasal samples (Roeder & Haas, 2013). In 2018, Albani and Fleming identified several new markers using RNA-Seq data from Lin et al.'s work from 2015. In this new study, several new markers were identified for blood, menstrual blood, semen, and more specifically spermatozoa (Albani & Fleming, 2018). The TNP1 transcript was found exclusively in semen samples containing sperm, while KLK2, which is specific for semen, was found in all semen samples. This differentiation of spermic and aspermic semen can allow for body fluids to be identified in individuals regardless of a person's reproductive ability. Not seen in the literature thus far are markers specific for urine, feces, breast milk, or tears. While these body fluids are not the most commonly found body fluids in evidence or at crime scenes, the ability to differentiate among all sources of genetic material shed from the human body is crucial.

Sequencing approach

With the advancements in next-generation sequencing, the analysis of RNA has not necessarily been a priority at all times. For this reason, the majority of NGS approaches still rely on sequencing complementary DNA (cDNA). The cDNA is synthesized through commercial Reverse Transcription (RT) kits that have varying RT enzyme properties that can affect the resulting product (Ståhlberg et al., 2004). This variation likely contributed to the differing assay sensitivities that were seen in the interlaboratory studies conducted by Ingold et al. in 2018 and 2020. Ideally, a reference marker would be included in all assays tested to monitor the effects of different RT kits, the presence of inhibitors, or normalization of data across samples due to a lack of mRNA-specific quantification methods. Unfortunately, the expression of published reference genes also referred to as housekeeping genes, has been shown to vary significantly between and even within individuals (Moreno et al., 2012). This can cause issues when comparing the results of different individuals. For this reason, the inclusion of a spike-in of synthetic RNA by the External RNA Controls Consortium is recommended to help normalize

the data (Pine et al., 2016). Unfortunately, this is typically done after extraction and RT of the sample and is therefore most helpful for gauging the efficacy of the library preparation process.

Library preparation of cDNA for RNA-seq follows a similar theme among the major commercial offerings. The cDNA can either be targeted with specific primer-adapter complexes or can be nonspecifically ligated with adapter sequences that are recognized by the respective sequencing platform. The sequencing platforms most commonly seen in the literature are offered by Illumina, as the MiSeq FGx, and ThermoFisher, as the Ion Torrent S5. These two instrumentation paths, while distinct, generally follow the same concept of sequencing by synthesis (SBS) following clonal amplification of captured molecules of cDNA on a flow cell (Kumar et al., 2019). This approach has relatively low error rates but is limited to short reads (~150 bp). Another commonality of SBS instruments is the error rate introduced by base misincorporation events. This fact, combined with the potential base misincorporation events during reverse transcription, means that RNA-seq data is particularly prone to having errors in the resulting data sets, causing downstream data analysis headaches (Le et al., 2013). An alternative to SBS exists in the nanopore sequencing approach employed by Oxford Nanopore Technology. Briefly, a motor protein pulls a nucleic acid molecule through a pore, and the alterations in the ionic current through the pore can determine the specific nucleotide order of the molecule. This sequencing method, combined with the newly launched Direct RNA Sequencing kit, which directly targets PolyA RNA, could eliminate the error introduced by reverse transcription and SBS (Depledge et al., 2019). At this time, however, the innate error rate for this technology is significantly higher than SBS, though improvements are being made at a rapid pace (Sahlin et al., 2021).

Beyond just body fluids

The sequencing of mRNA offers a wealth of information beyond just body fluid identification. And while outside the scope of this chapter, it should be noted that there are more applications. As previously discussed, the use of cSNPs in body fluid-specific mRNA markers can allow for the body fluids to be matched to individuals (Ingold et al., 2020). A set of mRNA markers was identified for correlation with the postmortem interval (Zhu et al., 2017). Advancements in this work could allow for the determination of time since death for individuals beyond the 27-h constraint of this particular study. The time since deposition of biological material can also be predicted for some of the commonly encountered body fluids in crime scenes up to 9 months old (Salzmann et al., 2019). A panel of 16 RNA markers was found to predict the chronological age of an individual to within 4 years in dermal fibroblast cells (Fleischer et al., 2018). While

variation would be expected in other body fluids, the ability to predict the age of donors from RNA-seq would be invaluable to investigators. While not mRNA markers, the long noncoding RNAs RPS4Y and XIST have been identified as gender-determining markers for men and women, respectively, and the markers are found in a variety of commonly encountered body fluids as they are generally considered housekeeping genes (Wang et al., 2021). A side product of RNA-seq analysis of body fluids is the inclusion of microbial species that have been coextracted. These microbial species have been found to be specific for different locations on the human body and are therefore correlated with specific body fluid samples (Salzmann et al., 2019; Swanson et al., 2020).

Analysis methods for mRNA data

As the adoption of RNA-Seq has increased, so have the methods for data analysis of the resulting sequence data. The standard treatment of data includes read trimming, quality control, alignment, and counting (Sheng et al., 2016). A large number of bioinformatic methods for the interrogation of gene and isoform abundance in datasets have emerged but are generally categorized into two groups: transcript-based and union-exon-based approaches. Transcript-based approaches attempt to analyze data by matching reads to individual isoforms (Li & Dewey, 2011). As some genes can have a large number of isoforms, accurately identifying a transcript read to a specific isoform is challenging. In union-exon methods, all exons of a gene are combined into a single union exon, and a transcript read is matched if it has sufficient identity with the union exon (Liao et al., 2014). In this way, reads can be matched to a gene much more easily, but isoform abundance in a sample is lost (Zhao et al., 2015). These methods allow for intersample differential gene expression to be determined for comparison. In Hanson et al.'s 2018 study, they employed the presence of mRNA markers in a sample to identify each of the body fluids in the study by agglomerative hierarchical cluster analysis (Hanson et al., 2018). This method treats the mRNA markers as a binary event that isn't prepared to handle mixtures of samples and would have difficulties with low-abundance samples missing some but not all markers characteristic of body fluid. Dørum et al. (2018), took the Hanson study data set and used the actual read counts of each gene to create a model based on partial least squares and linear discriminant analysis (Dørum et al., 2018). This approach resulted in higher predictive accuracy for the samples as well as the ability to accurately identify the different components in a body fluid mixture sample. To bring mRNA body fluid identification in line with contemporary DNA analysis, efforts have been made to apply likelihood ratios (LR) for body fluid prediction. Ypma et al. used in silico mixing of real body fluid mRNA profiles to test a prediction model based on the ratio H_1: the sample contains at least one body fluid of interest and H_2: the sample contains none of the

body fluids of interest (Ypma et al., 2021). The authors caution, however, that the resulting LRs won't be on the same scale as LRs generated for DNA profiles. This is because the LR for this model should be constrained to the number of data points in the H_2 category. To improve the results, a greater amount of data needs to be generated and compiled into a database. Additionally, efforts have been made to apply different machine learning methods to the analysis of RNA-Seq data rather than traditional differential gene expression (Wang et al., 2018). This approach uses artificial intelligence and information theory to make algorithms that learn from one dataset and then make predictions on new datasets.

Future of mRNA body fluid identification

The shift from serology testing in forensic laboratories is a necessary change in the world of forensic biology. An accurate confirmatory test derived from the same extraction that produces DNA profiles will allow examiners to confidently speak as to the origin of biological materials in a more precise manner. Next-generation sequencing of mRNA for body fluid identification offers exactly that confirmatory test. However, its implementation into the casework sections isn't ready yet. There are still a number of body fluids without known mRNA markers that haven't been reported in the literature as having been tested with the known markers of commonly encountered body fluids. To be able to confidently identify one body fluid requires the exclusion of all other body fluids. Urine, feces, tears, and breast milk remain the body fluids that, while not the most commonly submitted articles for evidence, do not yet have mRNA markers for confirmation. This must be rectified.

Additionally, the data that has been collected for mRNA sequencing thus far should be shared in a publicly accessible database. As there are a multitude of extraction methods, reverse transcription enzymes, library preparation processes, sequencing platforms, and bioinformatic pipelines, it is expected that there will be some variation in the data that is observed for a particular body fluid. A large collaboration and public sharing of data will allow researchers to better define the characteristics needed to identify body fluids in a manner that will be transparent and reliable. This transparent and reliable method of analysis could even come as a recommendation of specific or general guidelines as put forward by the regulatory agencies of the respective countries.

And finally, the most critical step for the introduction of next-generation sequencing of mRNA for body fluid identification to a wide audience of criminal investigation laboratories is testing the science in a court of law. Many laboratories around the world are hesitant to try new technologies—caution is absolutely warranted— to the extent that new methods languish in academic settings for years without making it to a courtroom. A crime lab needs to take it upon itself to forge a new path for body fluid identification and bring mRNA sequencing into the light.

References

Albani, P. P., & Fleming, R. (2018). Novel messenger RNAs for body fluid identification. *Science & Justice, 58*(2), 145–152. https://doi.org/10.1016/j.scijus.2017.09.002

Alvarez, M., Juusola, J., & Ballantyne, J. (2004). An mRNA and DNA co-isolation method for forensic casework samples. *Analytical Biochemistry, 335*(2), 289–298. https://doi.org/10.1016/j.ab.2004.09.002

Antunes, J., Silva, D. S. B. S., Balamurugan, K., Duncan, G., Alho, C. S., & McCord, B. (2016). Forensic discrimination of vaginal epithelia by DNA methylation analysis through pyrosequencing. *Electrophoresis, 37*(21), 2751–2758. https://doi.org/10.1002/elps.201600037

Bauer, M., Gramlich, I., Polzin, S., & Patzelt, D. (2003). Quantification of mRNA degradation as possible indicator of postmortem interval - a pilot study. *Legal Medicine, 5*(4), 220–227. https://doi.org/10.1016/j.legalmed.2003.08.001

van den Berge, M., Carracedo, A., Gomes, I., Graham, E. A. M., Haas, C., Hjort, B., Hoff-Olsen, P., Maroñas, O., Mevåg, B., Morling, N., Niederstätter, H., Parson, W., Schneider, P. M., Court, D. S., Vidaki, A., & Sijen, T. (2014). A collaborative European exercise on mRNA-based body fluid/skin typing and interpretation of DNA and RNA results. *Forensic Science International: Genetics, 10*(1), 40–48. https://doi.org/10.1016/j.fsigen.2014.01.006

Børsting, C., & Morling, N. (2015). Next generation sequencing and its applications in forensic genetics. *Forensic Science International: Genetics, 18*, 78–89. https://doi.org/10.1016/j.fsigen.2015.02.002

Bowden, A., Fleming, R., & Harbison, S. (2011). A method for DNA and RNA co-extraction for use on forensic samples using the Promega DNA IQ™ system. *Forensic Science International: Genetics, 5*(1), 64–68. https://doi.org/10.1016/j.fsigen.2009.11.007

Bustin, S. A. (2000). Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. *Journal of Molecular Endocrinology, 25*(2), 169–193. https://doi.org/10.1677/jme.0.0250169

Chirnside, O., Lemalu, A., & Fleming, R. (2020). Identification of nasal mucosa markers for forensic mRNA body fluid determination. *Forensic Science International: Genetics, 48*, 102317. https://doi.org/10.1016/j.fsigen.2020.102317

Depledge, D. P., Srinivas, K. P., Sadaoka, T., Bready, D., Mori, Y., Placantonakis, D. G., Mohr, I., & Wilson, A. C. (2019). Direct RNA sequencing on nanopore arrays redefines the transcriptional complexity of a viral pathogen. *Nature Communications, 10*(1). https://doi.org/10.1038/s41467-019-08734-9

Dørum, G., Ingold, S., Hanson, E., Ballantyne, J., Snipen, L., & Haas, C. (2018). Predicting the origin of stains from next generation sequencing mRNA data. *Forensic Science International: Genetics, 34*, 37–48. https://doi.org/10.1016/j.fsigen.2018.01.001

Fleischer, J. G., Schulte, R., Tsai, H. H., Tyagi, S., Ibarra, A., Shokhirev, M. N., Huang, L., Hetzer, M. W., & Navlakha, S. (2018). Predicting age from the transcriptome of human dermal fibroblasts. *Genome Biology, 19*(1). https://doi.org/10.1186/s13059-018-1599-6

Gauthier, Q. T., Cho, S., Carmel, J. H., & McCord, B. R. (2019). Development of a body fluid identification multiplex via DNA methylation analysis. *Electrophoresis, 40*(18–19), 2565–2574. https://doi.org/10.1002/elps.201900118

Haas, C., Klesser, B., Maake, C., Bär, W., & Kratzer, A. (2009). mRNA profiling for body fluid identification by reverse transcription endpoint PCR and realtime PCR. *Forensic Science International: Genetics, 3*(2), 80–88. https://doi.org/10.1016/j.fsigen.2008.11.003

Hanson, E., Ingold, S., Dorum, G., Haas, C., Lagace, R., & Ballantyne, J. (2019). Assigning forensic body fluids to DNA donors in mixed samples by targeted RNA/DNA deep sequencing of coding region SNPs using ion torrent technology. *Forensic Science International: Genetics Supplement Series, 7*(1), 23–24. https://doi.org/10.1016/j.fsigss.2019.09.011

Hanson, E., Ingold, S., Haas, C., & Ballantyne, J. (2018). Messenger RNA biomarker signatures for forensic body fluid identification revealed by targeted RNA sequencing. *Forensic Science International: Genetics, 34*, 206–221. https://doi.org/10.1016/j.fsigen.2018.02.020

Ingold, S., Dørum, G., Hanson, E., Ballard, D., Berti, A., Gettings, K. B., Giangasparo, F., Kampmann, M.-L., Laurent, F.-X., Morling, N., Parson, W., Steffen, C. R., Ulus, A., van den Berge, M., van der Gaag, K. J., Verdoliva, V., Xavier, C., Ballantyne, J., & Haas, C. (2020). Body fluid identification

and assignment to donors using a targeted mRNA massively parallel sequencing approach – results of a second EUROFORGEN/EDNAP collaborative exercise. *Forensic Science International: Genetics, 45*, 102208. https://doi.org/10.1016/j.fsigen.2019.102208

Ingold, S., Dørum, G., Hanson, E., Berti, A., Branicki, W., Brito, P., Elsmore, P., Gettings, K. B., Giangasparo, F., Gross, T. E., Hansen, S., Hanssen, E. N., Kampmann, M. L., Kayser, M., Laurent, F. X., Morling, N., Mosquera-Miguel, A., Parson, W., Phillips, C., … Haas, C. (2018). Body fluid identification using a targeted mRNA massively parallel sequencing approach – results of a EUROFORGEN/EDNAP collaborative exercise. *Forensic Science International: Genetics, 34*, 105–115. https://doi.org/10.1016/j.fsigen.2018.01.002

Juusola, J., & Ballantyne, J. (2007). mRNA profiling for body fluid identification by multiplex quantitative RT-PCR. *Journal of Forensic Sciences, 52*(6), 1252. https://doi.org/10.1111/j.1556-4029.2007.00550.x

Karlsson, M., Zhang, C., Méar, L., Zhong, W., Digre, A., Katona, B., Sjöstedt, E., Butler, L., Odeberg, J., Dusart, P., Edfors, F., Oksvold, P., von Feilitzen, K., Zwahlen, M., Arif, M., Altay, O., Li, X., Ozcan, M., Mardinoglu, A., … Lindskog, C. (2021). A single–cell type transcriptomics map of human tissues. *Science Advances, 7*(31). https://doi.org/10.1126/sciadv.abh2169

Kulstein, G., & Wiegand, P. (2018). Comprehensive examination of conventional and innovative body fluid identification approaches and DNA profiling of laundered blood- and saliva-stained pieces of cloths. *International Journal of Legal Medicine, 132*(1), 67–81. https://doi.org/10.1007/s00414-017-1691-6

Kumar, K. R., Cowley, M. J., & Davis, R. L. (2019). Next-generation sequencing and emerging technologies. *Seminars in Thrombosis and Hemostasis, 45*(7), 661–673. https://doi.org/10.1055/s-0039-1688446

Le, H. S., Schulz, M. H., Mccauley, B. M., Hinman, V. F., & Bar-Joseph, Z. (2013). Probabilistic error correction for RNA sequencing. *Nucleic Acids Research, 41*(10), e109. https://doi.org/10.1093/nar/gkt215

Liao, Y., Smyth, G. K., & Shi, W. (2014). FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics, 30*(7), 923–930. https://doi.org/10.1093/bioinformatics/btt656

Li, B., & Dewey, C. N. (2011). RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics, 12*. https://doi.org/10.1186/1471-2105-12-323

Lin, M. H., Albani, P. P., & Fleming, R. (2016). Degraded RNA transcript stable regions (StaRs) as targets for enhanced forensic RNA body fluid identification. *Forensic Science International: Genetics, 20*, 61–70. https://doi.org/10.1016/j.fsigen.2015.09.012

Lindenbergh, A., De Pagter, M., Ramdayal, G., Visser, M., Zubakov, D., Kayser, M., & Sijen, T. (2012). A multiplex (m)RNA-profiling system for the forensic identification of body fluids and contact traces. *Forensic Science International: Genetics, 6*(5), 565–577. https://doi.org/10.1016/j.fsigen.2012.01.009

Lin, M. H., Jones, D. F., & Fleming, R. (2015). Transcriptomic analysis of degraded forensic body fluids. *Forensic Science International: Genetics, 17*, 35–42. https://doi.org/10.1016/j.fsigen.2015.03.005

Madi, T., Balamurugan, K., Bombardi, R., Duncan, G., & Mccord, B. (2012). The determination of tissue-specific DNA methylation patterns in forensic biofluids using bisulfite modification and pyrosequencing. *Electrophoresis, 33*(12), 1736–1745. https://doi.org/10.1002/elps.201100711

Melé, M., Ferreira, P. G., Reverter, F., DeLuca, D. S., Monlong, J., Sammeth, M., Young, T. R., Goldmann, J. M., Pervouchine, D. D., Sullivan, T. J., Johnson, R., Segrè, A. V., Djebali, S., Niarchou, A., Wright, F. A., Lappalainen, T., Calvo, M., Getz, G., Dermitzakis, E. T., … Guigó, R. (2015). The human transcriptome across tissues and individuals. *Science, 348*(6235), 660–665. https://doi.org/10.1126/science.aaa0355

Moreno, L. I., Tate, C. M., Knott, E. L., Mcdaniel, J. E., Rogers, S. S., Koons, B. W., Kavlick, M. F., Craig, R. L., & Robertson, J. M. (2012). Determination of an effective housekeeping gene for the quantification of mRNA for forensic applications. *Journal of Forensic Sciences, 57*(4), 1051–1058. https://doi.org/10.1111/j.1556-4029.2012.02086.x

Old, J., Schweers, B. A., Boonlayangoor, P. W., Fischer, B., Miller, K. W. P., & Reich, K. (2012). Developmental validation of RSID^{TM}-semen: A lateral flow immunochromatographic strip test for the

forensic detection of human semen. *Journal of Forensic Sciences, 57*(2), 489–499. https://doi.org/10.1111/j.1556-4029.2011.01968.x

Old, J. B., Schweers, B. A., Boonlayangoor, P. W., & Reich, K. A. (2009). Developmental validation of RSID^{TM}-saliva: A lateral flow immunochromatographic strip test for the forensic detection of saliva. *Journal of Forensic Sciences, 54*(4), 866–873. https://doi.org/10.1111/j.1556-4029.2009.01055.x

Parker, C., Hanson, E., & Ballantyne, J. (2011). Optimization of dried stain co-extraction methods for efficient recovery of high quality DNA and RNA for forensic analysis. *Forensic Science International: Genetics Supplement Series, 3*(1), e309–e310. https://doi.org/10.1016/j.fsigss.2011.09.017

Park, J. L., Park, S. M., Kwon, O. H., Lee, H.c., Kim, J.y., Seok, H. H., Lee, W. S., Lee, S. H., Kim, Y. S., Woo, K. M., & Kim, S. Y. (2014). Microarray screening and qRT-PCR evaluation of microRNA markers for forensic body fluid identification. *Electrophoresis, 35*(21–22), 3062–3068. https://doi.org/10.1002/elps.201400075

Peters, M. J., Joehanes, R., Pilling, L. C., Schurmann, C., Conneely, K. N., Powell, J., Reinmaa, E., Sutphin, G. L., Zhernakova, A., Schramm, K., Wilson, Y. A., Kobes, S., Tukiainen, T., Ramos, Y. F., Göring, H. H. H., Fornage, M., Liu, Y., Gharib, S. A., Stranger, B. E., … Singleton, A. B. (2015). The transcriptional landscape of age in human peripheral blood. *Nature Communications, 6*. https://doi.org/10.1038/ncomms9570

Pine, P. S., Munro, S. A., Parsons, J. R., McDaniel, J., Lucas, A. B., Lozach, J., Myers, T. G., Su, Q., Jacobs-Helber, S. M., & Salit, M. (2016). Evaluation of the External RNA Controls Consortium (ERCC) reference material using a modified latin square design. *BMC Biotechnology, 16*(1). https://doi.org/10.1186/s12896-016-0281-x

Roeder, A. D., & Haas, C. (2013). MRNA profiling using a minimum of five mRNA markers per body fluid and a novel scoring method for body fluid identification. *International Journal of Legal Medicine, 127*(4), 707–721. https://doi.org/10.1007/s00414-012-0794-3

Sahlin, K., Sipos, B., James, P. L., & Medvedev, P. (2021). Error correction enables use of Oxford Nanopore technology for reference-free transcriptome analysis. *Nature Communications, 12*(1). https://doi.org/10.1038/s41467-020-20340-8

Sakurada, K., Akutsu, T., Watanabe, K., Fujinami, Y., & Yoshino, M. (2011). Expression of statherin mRNA and protein in nasal and vaginal secretions. *Legal Medicine, 13*(6), 309–313. https://doi.org/10.1016/j.legalmed.2011.07.002

Salzberg, S. L. (2018). Open questions: How many genes do we have? *BMC Biology, 16*(1). https://doi.org/10.1186/s12915-018-0564-x

Salzmann, A. P., Russo, G., Aluri, S., & Haas, C. (2019). Transcription and microbial profiling of body fluids using a massively parallel sequencing approach. *Forensic Science International: Genetics, 43*, 102149. https://doi.org/10.1016/j.fsigen.2019.102149

Scientific working group on DNA analysis methods (SWGDAM).(2015).

Setzer, M., Juusola, J., & Ballantyne, J. (2008). Recovery and stability of RNA in vaginal swabs and blood, semen, and saliva stains. *Journal of Forensic Sciences, 53*(2), 296–305. https://doi.org/10.1111/j.1556-4029.2007.00652.x

Sheng, Q., Vickers, K., Zhao, S., Wang, J., Samuels, D. C., Koues, O., Shyr, Y., & Guo, Y. (2016). Multi-perspective quality control of Illumina RNA sequencing data analysis. *Briefings in Functional Genomics, 16*(4), elw035. https://doi.org/10.1093/bfgp/elw035

Ståhlberg, A., Kubista, M., & Pfaffl, M. (2004). Comparison of reverse transcriptases in gene expression analysis. *Clinical Chemistry, 50*(9), 1678–1680. https://doi.org/10.1373/clinchem.2004.035469

Swanson, G. M., Moskovtsev, S., Librach, C., Pilsner, J. R., Goodrich, R., & Krawetz, S. A. (2020). What human sperm RNA-Seq tells us about the microbiome. *Journal of Assisted Reproduction and Genetics, 37*(2), 359–368. https://doi.org/10.1007/s10815-019-01672-x

Uhlén, M., Fagerberg, L., Hallström, B. M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, Å., Kampf, C., Sjöstedt, E., Asplund, A., Olsson, I. M., Edlund, K., Lundberg, E., Navani, S., Szigyarto, C. A. K., Odeberg, J., Djureinovic, D., Takanen, J. O., Hober, S., … Pontén, F. (2015). Tissue-based map of the human proteome. *Science, 347*(6220). https://doi.org/10.1126/science.1260419

Wang, T., Liu, Y., Ruan, J., Dong, X., Wang, Y., & Peng, J. (2021). A pipeline for RNA-seq based eQTL analysis with automated quality control procedures. *BMC Bioinformatics, 22*. https://doi.org/10.1186/s12859-021-04307-0

Wang, L., Xi, Y., Sung, S., & Qiao, H. (2018). RNA-seq assistant: Machine learning based methods to identify more transcriptional regulated genes. *BMC Genomics, 19*(1). https://doi.org/10.1186/s12864-018-4932-2

Watanabe, K., & Akutsu, T. (2020). Evaluation of a co-extraction kit for mRNA, miRNA and DNA methylation-based body fluid identification. *Legal Medicine, 42*, 101630. https://doi.org/10.1016/j.legalmed.2019.101630

Weinbrecht, K. D., Fu, J., Payton, M., & Allen, R. (2017). Time-dependent loss of mRNA transcripts from forensic stains. *Research and Reports in Forensic Medical Science, 7*, 1–12. https://doi.org/10.2147/RRFMS.S125782

Ypma, R. J. F., Maaskant - van Wijk, P. A., Gill, R., Sjerps, M., & van den Berge, M. (2021). Calculating LRs for presence of body fluids from mRNA assay data in mixtures. *Forensic Science International: Genetics, 52*, 102455. https://doi.org/10.1016/j.fsigen.2020.102455

Zhao, S., Xi, L., Zhang, B., & Jordan, I. K. (2015). Union exon based approach for RNA-seq gene quantification: To be or not to be? *PLoS One, 10*(11), e0141910. https://doi.org/10.1371/journal.pone.0141910

Zhu, Y., Wang, L., Yin, Y., & Yang, E. (2017). Systematic analysis of gene expression patterns associated with postmortem interval in human tissues. *Scientific Reports, 7*(1). https://doi.org/10.1038/s41598-017-05882-0

Zubakov, D., Boersma, A. W. M., Choi, Y., Van Kuijk, P. F., Wiemer, E. A. C., & Kayser, M. (2010). MicroRNA markers for forensic body fluid identification obtained from microarray screening and quantitative RT-PCR confirmation. *International Journal of Legal Medicine, 124*(3), 217–226. https://doi.org/10.1007/s00414-009-0402-3

Zubakov, D., Kokmeijer, I., Ralf, A., Rajagopalan, N., Calandro, L., Wootton, S., Langit, R., Chang, C., Lagace, R., & Kayser, M. (2015). Towards simultaneous individual and tissue identification: A proof-of-principle study on parallel sequencing of STRs, amelogenin, and mRNAs with the ion torrent PGM. *Forensic Science International: Genetics, 17*, 122–128. https://doi.org/10.1016/j.fsigen.2015.04.002

CHAPTER 22

NGS-based detection and differentiation of forensically relevant body fluids using conventional, molecular, and microbial techniques

Hirak Ranjan Dash
Department of Forensic Science, National Forensic Sciences University, New Delhi, India

Background

The detection of various body fluids such as blood, semen, vaginal fluid, saliva, menstrual blood, and urine is a key aspect of the criminal investigation. Identification of body fluid type and origin helps in the reconstruction of a crime scene by linking the sample donors and real criminal acts (An et al., 2012). Many body fluids are invisible to the naked eye, and most of them share a common morphology, which makes it difficult to distinguish them (Harbison & Fleming, 2016). Besides, when two body fluids are mixed at a crime scene, it becomes immensely strenuous for the investigators to decipher the nature of the body fluid. In current practice, the body fluids/stains are ultimately used for individualization through DNA profiling. However, the generation of a DNA profile from the wrongdoer does not disclose the situations by which it was transferred. Correct identification of body fluids before DNA profiling provides a source for a DNA profile from the samples, which ultimately increases its evidentiary value as well as providing investigative leads in a case.

For better identification of body fluids, innovative techniques are being developed with little consumption of evidence. In this regard, attenuated total reflection Fourier transform infrared (ATR-FTIR) spectroscopy and Raman spectroscopy coupled with chemometrics have shown huge potential as nondestructive body fluid identification techniques (Elkins, 2011; Takamura et al., 2018; Virkler & Lednev, 2010). The detection of tissue-specific expression profiles of microRNAs (*miRNAs*) using quantitative reverse transcription PCR (RT-PCR) has also been used as an alternative molecular body fluid identification technique (Zubakov et al., 2010). Besides, different RNA markers and tissue-specific differentially methylated regions have been identified to distinguish forensically relevant body fluids (Lee et al., 2012). For rapid detection of blood, semen, and/or saliva at the crime scene, quantum dot molecular beacon-based techniques have also been developed to test tissue-specific RNAs (Young et al., 2017). Additionally, many biosensor devices have been developed for real-time analyte/body fluid detection using

Next Generation Sequencing (NGS) Technology in DNA Analysis
ISBN 978-0-323-99144-5, https://doi.org/10.1016/B978-0-323-99144-5.00003-2
© 2024 Elsevier Inc.
All rights reserved.

innovative optical sensing techniques (Frascione et al., 2013). Other techniques, such as immunochromatographic strip tests, have also been developed for onsite body fluid detection (Old et al., 2012).

Identification of body fluid alone does not necessarily meet the purpose of forensic analysis. Species determination is also of utmost importance for providing investigative leads. In most cases, limited samples are available for forensic analysis. The preliminary tests being conducted for body fluid screening are highly destructive, giving rise to sample-limiting conditions for further individualization testing. Rapid on-site testing of body fluids without waiting for laboratory-generated reports is highly useful for forensic investigation purposes. Though significant advancements have taken place in body fluid identification research and many products are available for their detection, a practical model with a gold-standard application is still lacking (Virkler & Lednev, 2009). Thus, a body fluid detection technique with no destructiveness, little consumption of samples, a successful outcome on a variety of substrates, and high sensitivity has become the need of the hour. An overview of the technologies used for forensic investigation is presented in Fig. 22.1.

The human body acts as a micro-ecology and harbors indigenous microbiota as commensals (Tannock & Tannock, 1999). Extensive analysis of human normal microflora reveals the presence of predominant species types in different body niches such as the oral cavity, skin, vagina, stomach, large and small intestine, and urinary tract (McFarland, 2000). In this regard, analysis of microbial patterns shows huge potential as a better alternative for forensically relevant body fluid detection (Hanssen et al., 2017; Oliveira & Amorim, 2018). Though much progress has taken place in microbial-based human body fluid identification research to date; however, the development of such a model is still at a nascent stage. The current chapter emphasizes the availability of current

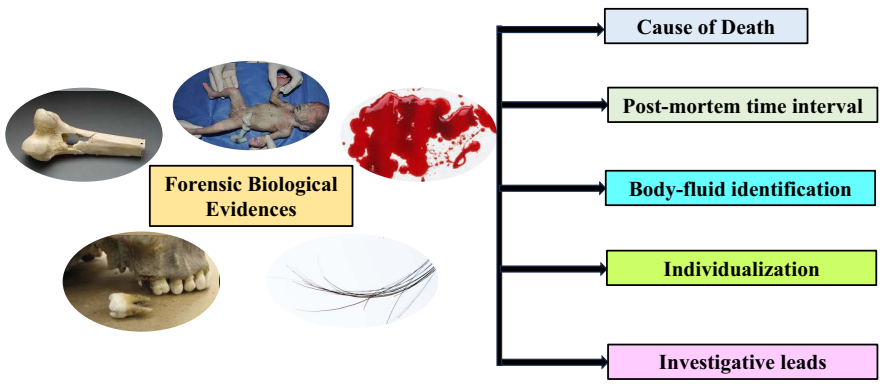


Figure 22.1 Various aspects of criminal investigations using body fluids found at the crime scene.

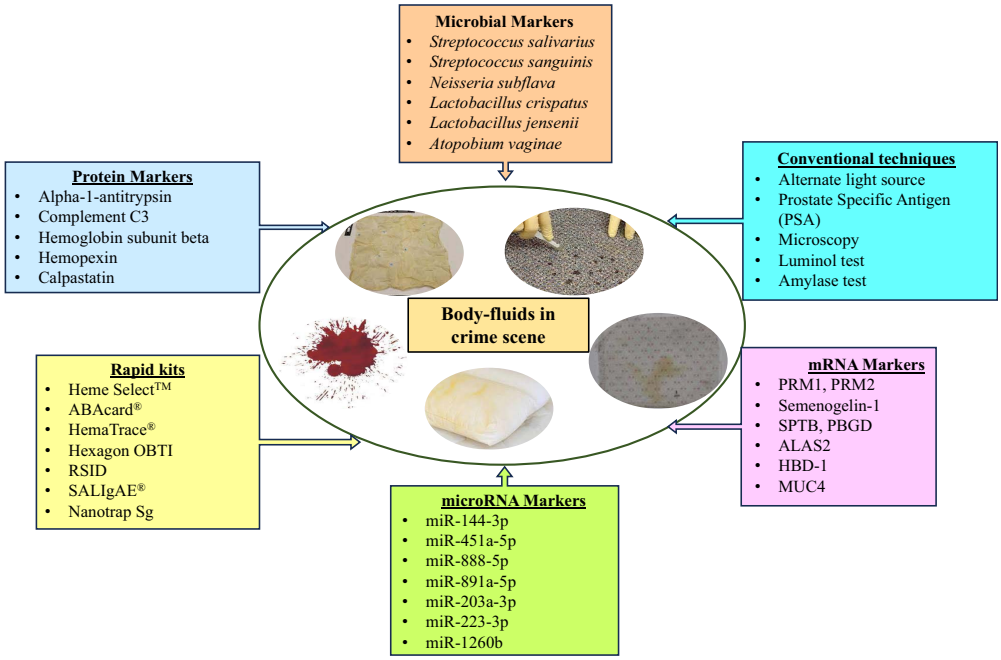


Figure 22.2 Use of various techniques and molecular markers for identification of body fluids from biological samples.

conventional techniques for human body fluid detection and the exploitation of other technological advancements in the field of body fluid identification research, discussing their respective pros and cons. Different techniques used for the identification of body fluids have been summarized in Fig. 22.2.

Existing technologies for human body fluid detection

Many body fluid identification techniques have been developed over the past century and are currently in practice in different forensic science laboratories throughout the world. Most of them include chemical tests, immunological tests, protein catalytic activity tests, spectroscopic methods, and microscopy (An et al., 2012), which have been described as follows: However, most of them are presumptive, and the search for a gold standard technique for human-specific body fluid identification continues.

Conventional techniques

Semen

The detection of human semen is one of the most prerequisites in investigation of sexual assault cases. The most common techniques currently used for human semen detection

include a visual examination, followed by the use of ultraviolet light to detect fluorescent properties of semen and chemical analysis for the detection of acid phosphatase, an important component of seminal fluid (Stefanidou et al., 2010). On manual observation, the seminal fluid on any fabric appears viscous, stiff, and crusty. For further confirmation, they should be observed under UV light between 400 and 700 nm, as semen is fluorescent due to the presence of conjugated chlorine and/or flavin proteins (Kobus et al., 2002). However, the presence of ointments and creams on fabric has been reported to give false-positive results using alternate light sources (ALS) (Santucci et al., 1999). In this regard, other commercial ALS, such as Bluemaxx BM500 and Polilight, have shown better sensitivity toward seminal fluid detection. But none of these light sources predicts the human specificity of the seminal fluid. Further, the detection of seminal acid phosphatase is widely used for the recognition of seminal fluid. After hydrolysis of organic phosphates by the acid phosphatase enzyme, the hydrolyzed product reacts with diazonium salt chromogen to give rise to a change in color. A combination of many compounds, such as alpha-naphthyl phosphate and Brentamine Fast Blue, beta-naphthol and Fast Garnet B, alpha-naphthol and Fast Red AL, p-nitroaniline, $NaNO_3$, a-naphthyl phosphoric acid, and aqueous magnesium chloride, is popularly in practice for this test (Banerjee, 1983; Gaensslen, 1983; Kosa et al., 1987).

However, any of the previously described tests cannot be considered confirmatory as studies have shown false positive results with certain plant materials such as strawberries, broad beans, onions, plums, vaginal acid phosphatase, and feces (Gaensslen, 1983; Kosa et al., 1987; Vennemann et al., 2014), and acid phosphatase tends to degrade over time due to exposure to heat, mold, putrefaction, or chemicals (Takayama et al., 2001). The most widely accepted confirmatory test for semen detection is the microscopic identification of sperm cells using Gram-modified Christmas tree stain, hematoxylin and eosin, Baecchi's, Papanicolaou's, and Wright's stains (Kondracki et al., 2017). In conditions of natural azoospermic individuals and vasectomized persons, the microscopic examination does give a false negative result. Thus, other techniques have been developed to detect prostate-specific antigen, seminal vessel-specific antigen, semenogelin (Sg), Sg II, 19-OH $F_{1\alpha}/F_{2\alpha}$ prostaglandin, and monoclonal antibody 1E5 with limited cross-reactivity (Virkler & Lednev, 2009). Each of these techniques has its own pros and cons. The user has to choose the appropriate methodology for the detection of semen as a body fluid as per its requirements.

Blood

Blood and bloodstains are the most common biological evidence found at a crime scene. To detect latent bloodstains at crime scenes and/or on a dark background, UV light as ALS can be applied using light source products such as Polilight. A recent study showed that UV light at 365 nm wavelength is highly effective in detecting even traces of human

blood serum (Kearse, 2020). However, a comparative analysis of UVA, UVB, and UVC showed that UVC is the most damaging to human blood DNA, as many STR loci could not be amplified after exposure to UVC (Uchigasaki et al., 2018). Another study showed the presence of bipyrimidine photoproducts, oxidative lesions, and strand breaks followed by inhibition of STR profiling when bloodstains are exposed to UV light of 25 to 1236 J/cm^2 (Hall et al., 2014). Thus, ALS particularly the shorter-wavelength UV light should be carefully used while detecting latent bloodstains on any surface. Another presumptive test for blood that has been carried out at the crime scene for a very long time is the luminol test. The principle behind this test is that hemoglobin and other blood derivatives enhance the oxidation of luminol in an alkaline environment. Popular forms of luminol testing such as Grodsky formulations, Weber, Weber II, and Bluestar1 alternatives have also shown detrimental effects on subsequent DNA analysis besides giving false positive and/or negative results in surfaces covered with terracotta, ceramic tiles, polyurethane varnishes, or jute and sisal matting (Passi et al., 2012; Quickenden & Creamer, 2001). Another presumptive test, the benzidine test, which exploits the peroxidase-like activity of the heme group, has become obsolete due to the carcinogenic nature of benzidine and the cross-reactivity of this test with chemical oxidants and plant peroxidases (Colotelo et al., 2009). Other catalytic tests, such as the phenolphthalein test and leucomalachite green test are not as sensitive as the luminol test but can detect blood at a dilution of 1:10,000. However, such tests do not give the species origin of bloodstains.

A microscopic examination of red and white blood cells along with fibrin can be considered for confirmation of bloodstains. The most popular blood confirmation test is the Teichman crystal test, where hematin is formed when a dried bloodstain is heated in the presence of a halide and glacial acetic acid. Measurement of absorbance at 400 nm, the absorption maxima of hemoglobin can also confirm the body fluid. However, hindrance from water submersion, sunlight exposure, heating, and rust has been reported in this technique (Bremmer et al., 2011). Isozyme analysis to compare lactate dehydrogenase isozyme patterns and enzyme-linked immunosorbent assays are other confirmatory tests for the detection of human blood besides identifying different blood groups using specific antibodies (Virkler & Lednev, 2009). Besides, the chromatography technique for the separation of hemoglobin and its derivatives by the development of hematoporphyrin can also confirm the presence of blood. However, the lengthy development time and necessity of presaturation limit the use of this technology in current-day forensic practice.

Further, it becomes immensely difficult to distinguish between peripheral blood and menstrual blood. In this regard, a duplex test combining human hemoglobin and D-

dimer detection has recently been developed to detect blood and menstrual fluid (Holtkötter et al., 2018). Applications of ATR-FTIR spectroscopy and chemometrics have also shown a huge potential to distinguish between peripheral and menstrual blood (Sharma et al., 2020).

Vaginal fluid

Vaginal fluid as evidence plays a major role in sexual assault cases. As the composition and pH of vaginal fluid change over different phases of the menstrual cycle, testing specific components is highly challenging (Wagner & Levin, 1980). However, many presumptive and confirmatory tests have been proposed and are currently in use in forensic practice. The most common test includes the detection of glycogenated epithelial cells using a periodic acid-Schiff reagent. Shreds of evidence showed that women in the prepubescent or postmenopausal stages do not have glycogenated cells (Anderson et al., 2012), which tends to give false-negative results in those cases, questioning the reliability of this test. Other techniques to detect vaginal peptidase, esterase, alkaline phosphatase, b-glucuronidase, DPNH-diaphorase, and estrogen receptors have also been developed for the identification of vaginal fluid (Divall, 1984; Gaensslen, 1983; Hausmann et al., 1996). However, most of these techniques have their loopholes, as false-positive results have been reported from male prepuce and urethral mucosa samples. In another approach, pristine vaginal fluid can be distinguished from a mixture of vaginal and seminal fluid by analyzing the ratio of lactic acid and citric acid, as vaginal fluid contains large quantities of lactic acid whereas seminal fluid contains large quantities of citric acid (Aldunate et al., 2013). As both of these carboxylic acids are found in traces in other body fluids such as saliva and urine, such preliminary tests need to be confirmed by other molecular levels of examination.

Saliva

Besides semen and blood, saliva is also found commonly at a crime scene which needs to be detected by forensic examination. Though the many presumptive tests have been reported for this body fluid, no confirmatory test has been proposed to date for saliva detection. ALS is widely used to locate saliva stains as they appear bluish-white under UV (Fiedler et al., 2008). Though detection of activity of amylase is widely used as a presumptive test for saliva due to the presence of more amount of AMY1 in saliva, cross-reactivity of this test has been detected in other body fluids such as pancreatic secretions, semen, and vaginal fluid. Such tests also generate false-positive results for hand cream, face lotion, washing powders, urine, and feces (Virkler & Lednev, 2009). Microscopic examination such as scanning electron microscopy (SEM) coupled with energy dispersive X-ray (EDX) provides tentative identification of saliva by detecting the relative concentration of sodium, phosphorus, sulfur, chlorine, potassium, calcium, and other trace

metals, where potassium is found in the form of the largest peak (Saibaba et al., 2017). The use of such a sophisticated instrument for preliminary screening of body fluids can be a huge burden on the laboratory's resources.

Urine

In cases of sexual assault, harassment, and mischief, the examination of urine as a body fluid plays a great investigative lead. The unique odor of urine can be found on the whole object, but it becomes immensely difficult to localize the stain's position as it spreads out quickly and is diluted on any object. Ionic characterization of urine shows the presence of a higher amount of phosphate and sulfate, suggesting its usefulness in body fluid detection. The use of Nessler's reagent, or p-dimethylaminocinnamaldehyde, as an indicator for the indirect detection of urea present abundantly in human urine is also used successfully (Ong et al., 2012). In some cases, creatinine and Tamm-Horsfall glycoprotein (THP) are also estimated for urine detection; however, such tests can also give false-positive results for urine samples of other species. SEM-EDX, high-performance liquid chromatography, and electrospray ionization liquid chromatography-mass spectrometry (LC-MS) are widely used to confirm the detected human urine samples.

Sweat

Though sweat is found in rare instances in crimes, it shows huge potential as a source of DNA to be used in the individualization process. The composition of sweat and urine is nearly the same except containing a little amount of urea and creatinine. Identification of relative concentrations of sodium, phosphorus, sulfur, chlorine, potassium, calcium, and other metal ions using SEM-EDX can indicate the presence of sweat in a sample. Analysis of such samples shows a characteristically huge peak of chlorine in comparison to other ionic peaks, suggesting the sweat origin of this sample (Seta, 1977). However, this test is not confirmatory and does not give any clue about the species origin of the sample. Multidimensional spectroscopic analysis by Raman spectroscopy shows spectral components from lactate, lactic acid, urea, and single amino acids, suggesting the usefulness of this technique in sweat detection (Sikirzhytski et al., 2012).

Molecular techniques

Semen

The detection of cell-specific gene expression has shown huge potential for confirmatory identification of human seminal fluid. Protamines, exclusively expressed in the haploid genome during spermatogenesis, have been explored for the detection of dried semen stains (Bauer & Patzelt, 2003). Analysis of three semen-specific mRNA markers, i.e., Protamine-1 (PRM1), protamine-2 (PRM2), and semenogelin-1 (SEMG1), in 50 samples showed their exclusive detection in semen samples with no cross-reactivity with the

blood sample. Further, this study demonstrated the PRM1 marker to be the most reliable marker, followed by PRM2 and SEMG1, for human seminal fluid detection (Nader et al., 2015). In this regard, a multiplex mRNA profiling system has been proposed that includes semen-specific markers such as PRM2 (protamine 2), TGM4 (transglutaminase 4), and SEMG1, besides simultaneous detection of other human body fluids (Song et al., 2015). In a recent study, relative quantification of miRNAs such as miR-144-3p, miR-451a-5p, miR-888-5p, miR-891a-5p, miR-203a-3p, miR-223-3p, and miR-1260b was useful in discriminating four different types of body fluids (Fujimoto et al., 2019). Circular RNAs, the products of backsplicing of premRNAs, have also shown huge promise in human seminal fluid detection with markers such as TGM4, KLK3, and PRM2 (Liu et al., 2019). Analysis of methylome data on 16 samples showed cg23521140 and cg17610929 CpG sites to be novel DNA methylation markers for semen samples, suggesting these markers to be more effective in human seminal fluid detection (Park et al., 2014). In cold cases with limited availability of RNA and the non-suitability of DNA methylation analysis, the analysis of copy number variation (CNV) of certain tissue-specific DNA markers has also been studied for human body fluid detection. Analysis of CNV of mtDNA and telomerase repeat length has been reported to be most suitable for human seminal fluid detection through qRT-PCR (Zubakov et al., 2018). Genetic markers have shown huge potential in differentiating human seminal fluid from other body fluids; however, the examiner/analyst should undertake the technique based on their expertize and infrastructure availability.

Blood

Human immunoglobulin G (IgG) is the most common immunoglobulin in human blood and is present continually (Gonzalez-Quintela et al., 2008). Though, irrespective of the species variation, the structure of IgG is similar, the presence of certain species-specific epitopes mostly the glycosylation of IgG (Shantha Raju et al., 2000), makes them suitable target entities for the detection of human blood for forensic applications. Besides immunoglobulins, a variety of serum proteins, such as human HbA0, HLA antigens, and muscle-specific beta-enolase have been explored for human-specific detection of blood (Hurley et al., 2009).

Specific mRNA molecules can also be targeted in the evidentiary samples for specific body fluid detection, depending on the cell type and its function. β-spectrin and porphobilinogen deaminase genes and their respective mRNAs have been established to specifically target human blood (Juusola & Ballantyne, 2005). In another study, delta-aminolevulinate synthase 2 has been reported as a suitable mRNA marker for the detection of peripheral blood (Richard et al., 2012). In this regard, the matrix metalloproteinase-11 (MMP-11) gene, which codes for proteins involved in extracellular matrix breakdown, is highly useful in detecting menstrual blood. The MMP1 gene is found to be upregulated considerably at the time of menstruation in comparison to

the other phases of the menstrual cycle; hence, targeting its mRNA can conclusively distinguish between the peripheral blood and menstrual blood (Counsil & McKillip, 2010).

Protein biomarkers are one of the most promising targets for human body fluid detection. The advantage of protein biomarkers over other techniques is their diversity, stability over time, and tissue-specific posttranslational modifications. In this regard, alpha-1-antitrypsin, complement C3, hemoglobin subunit beta, and hemopexin have been explored for their suitability in human peripheral blood detection and calpastatin for human menstrual blood detection at an efficacy level of 96% (Legg et al., 2014). The 4%–12% cross-reactivity of these markers has been attributed to blood from flossing teeth, urinary tract infections, the presence of minor vaginal abrasions, or residual menstrual fluid in the vaginal canal. In another study, protein biomarkers glycophorin A and hemoglobin β have been found useful in detecting human blood by dot blot assay in aged and simulated forensic evidence with high specificity and sensitivity (de Beijer et al., 2018).

Vaginal fluid

For a forensic biological examiner, it becomes immensely difficult to distinguish vaginal fluid from other similar-looking body fluids such as nasal secretions, saliva, urine, semen, and sweat. Many biomarkers have been studied for the proper detection and identification of vaginal fluid as forensic evidence. Human small proline-rich protein 3 (SPRR3) isoforms and human fatty acid-binding protein 5 (FABP5) with an acetylated N-terminal region lacking the initiator methionine residue have been explored for high-fidelity detection of human vaginal fluid (Igoh et al., 2015). Multiplex analysis targeting ESR1, SERPINB13, KLK13, CYP2B7P1, and mucin 4 (MUC4) genes has been found useful for the identification of vaginal fluid (Akutsu et al., 2017, 2020). In a novel study, 17β-estradiol (E2-17β), SPRR3, and FABP5 have been targeted to determine human vaginal fluid by GC-MS and LC-MS (Igoh et al., 2015; Sakurada et al., 2008).

RT-PCR-based detection of human beta-defensin 1 and MUC4 markers in mRNA has been widely accepted for the identification of vaginal fluid (Cossu et al., 2009a). Besides, E2-17b, which has the strongest estrogenic activity of all the female hormones has been detected by GC-MS confirming the biological fluid as vaginal secretion. ELISA-based tests utilizing E2-17b monoclonal antibodies have also shown usefulness for this hormonal-based detection of human vaginal fluid (Virkler & Lednev, 2009).

Saliva

Saliva, the complex biological fluid secreted by acinar cells of the major and minor salivary glands plays an important forensic role. Besides individualization using saliva samples, it can also be used for other forensic applications such as an indicator of salivary gland conditions, toxicology, and drug monitoring (Chatterjee, 2019). As amylase is

one of the major constituents of saliva, most of the techniques explore the detection of α-amylase to identify this body fluid. However, most of these tests cannot differentiate between α-amylase 1 present in saliva, and α-amylase 2 present in semen and vaginal secretions. With the advent of technology, many molecular markers and techniques have been discovered to detect human saliva as a body fluid with increased forensic importance.

Many miRNA markers, such as miR-26a, miR-96, miR-135b, miR-138, miR-141, miR-145, and miR-182, have been explored for exclusive detection of human saliva in forensic samples (Mishra et al., 2021). In this regard, Silva et al. (2016) suggested BCAS4 as an important epigenetic marker for human saliva detection. Salivary RNA is being speculated as one of the most useful biomarkers for the detection of human saliva. Though the detection of salivary mRNA is an emerging field, both the RT-PCR technique and gel electrophoresis technique have been used for the detection of human saliva, targeting statherin and histatin 3 as indicators (Juusola & Ballantyne, 2005). Most of the salivary protein markers are used for the detection of diseases. However, they can be envisioned to detect this body fluid with forensic relevance. The most commonly used protein biomarkers for the detection of saliva include glyceraldehyde-3-phosphate dehydrogenase, beta-actin, and ribosomal protein S9 (Lee et al., 2011). Several other protein markers, such as alpha-amylase 1, mucin-5B, and proline-rich protein HaeIII subfamily 2 protein are also explored for the determination of saliva and saliva stains (de Beijer et al., 2018). Such molecular markers for the identification of human-specific saliva are still being explored for their suitable usability in forensic detection.

Urine

Urine is one of the majorly encountered biological fluids found at the crime scene. Besides the purpose of identification, urine is considered to be the most useful body fluid for toxicological and drug detection over blood or hair (Usman et al., 2019). Few markers provide huge genetic polymorphism in human urine than those in blood serum, such as uropepsinogen, DNase I, and DNase II. Considering the huge content of these compounds in urine, they can also be speculated as suitable markers to detect this body fluid (Yasuda et al., 1997). A study on nuclear DNA and mitochondrial DNA from urine samples showed a higher recovery of mitochondrial DNA even after 4 months of storage (Castella et al., 2006). Due to the superior usability of urine samples, detection of the origin of this body fluid becomes imperative from a forensic investigation point of view. In this regard, two RT-PCR-based genetic markers, i.e., renin and uroplakin 2, have been recommended as the two most suitable markers for human urine detection in addition to their huge diagnostic value (Matuszewski et al., 2018). TMP, a constituent present in human as well as other animals' urine has also been recommended as a suitable marker for urine detection (Harbison & Fleming, 2016). Several protein markers, such as

osteopontin and uromodulin are also speculated to be useful for the detection of human proteins using ELISA and immunochromatographic tests (Legg et al., 2014). Other epigenetic markers used for the diagnosis of prostate cancer, such as transmembrane protease serine-2 and ETS-related gene fusion, can also be used for the detection of urine (Dimitriadis et al., 2013).

Sweat

Dermcidin (DCD), a small antibiotic peptide secreted into human sweat is considered a suitable marker for the detection of sweat. Sakurada et al. (2010) evaluated the performance of both RT-PCR and ELISA assays for the detection of DCD and found that RT-PCR could detect sweat using DCD in 7-day-old stains containing as small as 10 μL of sweat. Besides, mucin-5B and cathepsin D have also shown their potential as useful markers for the detection of human sweat samples (de Beijer et al., 2018). Cystic fibrosis transmembrane conductance regulating (CFTR) proteins present in sweat glands are responsible for the transport of Na and Cl in epithelial-secreting cells. Modifications in this protein are responsible for cystic fibrosis disease determination (Jadoon et al., 2015). Thus, besides the detection of the disease, the CFTR marker can be speculated to be a useful marker for the detection of human sweat. Another sweat-specific protein marker, G-81, produced by the monoclonal antibody technique, has been found useful to specifically detect human sweat samples (Sagawa et al., 2003).

Rapid techniques

With the advancement in technology, many rapid tests for body fluid detection have been discovered over the years. In addition to the screening tests, some noncatalytic presumptive tests have been developed for different body fluids. For example, Heme Select, ABAcard, HemaTrace, and Hexagon OBTI are being used to identify primate blood. Out of all these rapid techniques, Hexagon OBTI has been found to perform well on aged and degraded blood samples (Spalding et al., 2003). The use of a lateral flow strip containing antiglycophorin A has been used recently for the rapid detection of human blood by identifying sialoglycoprotein (Reich, 2008). Similarly, a new commercially available rapid stain identification semen test has been discovered that targets semenogelin (Sg) in semen protein by using antihuman Sg antibodies. This technology of using immunochromatographic membrane assay technology is found to be more sensitive by detecting semen at 1:100,000 dilutions (Pang & Cheung, 2007). Nanotrap Sg was also recently developed as a rapid test kit for the detection of human semen stains with the same efficiency (Laffan et al., 2011; Sato et al., 2007).

Besides, many rapid tests have also been developed for the detection of human saliva. A lateral flow strip method using nine antibodies against human salivary amylase has been developed for the detection of human saliva. This technique has been found useful in

different challenging samples, such as buccal swabs, plastic bottles, plastic mugs, ceramic mugs, cigarette butts, and soda cans (Reich, 2007). Another advanced method available in the form of a spray is also available for colorimetric detection of saliva, called SALIgAE. Though this technique is much useful in field conditions, it was found to be less sensitive in comparison to other available rapid tests for human saliva (Li, 2008). Similarly, rapid testing techniques from Galantos Genetics are available for the detection of human urine from a variety of samples. Rapid detection of other body fluids, such as human-specific vaginal fluid and sweat, needs to be developed to have a complete array of rapid techniques for human-specific body fluid detection.

Microbial-based body fluid detection

A human body contains around 10^{14} microbial cells. Human body fluids were earlier considered to be sterile, whereas studies have also shown the presence of different microorganisms in different body fluids. Many commensal and pathogenic microbes commonly inhabit specific body parts, contributing to their presence in different body fluids. A list of signature microflora found in different body fluids is given in Table 22.1. The advantage of microbes is their abundance and stability. As different microorganisms are found in different body parts, a subset of genetic markers targeting the dominant microflora can be used to determine the body fluid.

With the advent of molecular biology techniques, the analysis of microflora DNA from different fluids has shown promising results in the detection of these body fluids (Giampaoli et al., 2012). A recent study showed the use of 16S rRNA and *rpoB* as target regions not only to detect saliva as the body fluid but also to distinguish between the saliva of two different individuals (Leake et al., 2016). Another study on 140 Korean individuals showed the suitability of the OB mRT-PCR technique targeting the combinatorial detection of three oral bacteria, *Streptococcus salivarius*, *Streptococcus sanguinis*, and *Neisseria subflava,* for the identification of human saliva (Jung et al., 2018). Similarly, analysis of the 16S rRNA gene of *Lactobacillus crispatus*, *Lactobacillus jensenii,* and *Atopobium vaginae* was found to be useful for the detection of human vaginal fluid (Akutsu et al., 2012). In this context, a microbial-based RT-PCR kit (ForFLUID kit) has been developed to discriminate among human vaginal, oral, and fecal samples (Giampaoli et al., 2014).

A taxonomy-independent deep neural network approach based on massively parallel microbiome sequencing using the 16S rRNA gene showed differentiation of human blood from four different body sites, such as menstrual, nasal, fingerprick, and venous blood. The level of accuracy of this technique for the detection of different body fluids, such as 0.992 for menstrual blood, 0.978 for nasal blood, 0.978 for fingerprick blood, and 0.990 for venous blood, showed the usefulness of microbial analysis for the detection of

Table 22.1 Signature microflora found on different tissues/body fluids for use in body fluid identification.

Sl. no.	Body fluids/ tissues	Signature microflora	Genetic markers for identification	References
1	Vaginal fluid	*Veillonella atypical*, *Lactobacillus crispatus*, *Lactobacillus gasseri*, *Atopobium vaginae*	Multiplex PCR of 16S rRNA gene, 16−23S rDNA	Akutsu et al. (2012), Fleming and Harbison (2010), Giampaoli et al. (2014), Jung et al. (2018)
2	Bloodstains	*Proteobacteria*, *Actinobacteria*, *Firmicutes*, *Bacteroidetes*	16S rRNA gene sequencing	Castillo et al. (2019)
3	Saliva	*Streptococcus salivarius*, *Streptococcus sanguinis*, and *Neisseria subflava*	16S rRNA and rpoB PCR, 16S−23S ITS, streptococcal-specific PCR-DGGE	Choi et al. (2014), Hsu et al. (2012), Jung et al. (2018), Leake et al. (2016)
4	Semen stains	*Staphylococcus* sp. *Corynebacterium* sp.	16S rRNA gene sequencing	Yao et al. (2020)
5	Menstrual blood	*Lactobacillus crispatus*, *Lactobacillus gasseri*	Multiplex PCR of 16S rRNA gene	Choi et al. (2014)
6	Sweat	*Corynebacterium* sp. *Staphylococcus* sp. *Propionibacterium* sp.	16S rRNA gene sequencing	Peterson et al. (2016)
7	Urine	*Firmicutes* sp. *Bacteroidetes* sp. *Actinobacteria* sp. *Fusobacteria* sp. *Proteobacteria* sp.	Culture based approach	Neugent et al. (2020)

blood as well as its origin to the part of the body (López et al., 2020). However, contradictory findings were reported by Dobay et al. (2019), as vaginal and menstrual blood share common microbial signatures, resulting in their indistinguishability using the microbial approach. Hence, though the microbial markers are a promising avenue for body-fluid detection/tissue differentiation, a thorough study is required before applying this technique in real-time forensic case solving.

Technological advancements in microbial forensics

Technological advancements in the microbial analysis of biological samples have reached the genomic and postgenomic eras. In this regard, most of the studies focus on the sequencing of 16S rRNA genes in bacteria. Studies involving the analysis of the microbiome in body fluids and tissue samples such as saliva, skin, peripheral blood, menstrual blood, feces, and semen have used this technique to achieve sufficient discrimination power to distinguish among tissue types (Sijen & Harbison, 2021). Many studies have advocated the use of a whole-genome sequencing (WGS) approach for highly accurate data in body fluid detection using the microbial approach; however, the per-sample analysis cost is a huge drawback to this approach (Ranjan et al., 2016).

Besides the sequencing of the 16S rRNA gene or whole genome, many other techniques are also being used for microbial-based body fluid detection. In this regard, microarray-based detection of 16S rRNA, GroEL, and 18S rRNA genes of several microorganisms present in feces and other forensically relevant fluids has opened a new era for the utilization of this advanced technique for the detection of human body fluids (Quaak et al., 2017). Application of the RT-PCR technique has also been found useful in the detection of body-fluid-specific microorganisms. RT-PCR-based detection of three oral bacteria with 10^2–10^7 copies/reaction was found useful in the detection of human saliva (Jung et al., 2018). With the advent of next-generation sequencing (NGS) technology, microbial-based body fluids detected are also deemed real using this most advanced technology (Yang et al., 2014).

Though detection of body fluids using microbial markers is a promising field, applications of microbial markers in tracing geolocation, personal identification, determination of biological sex, manner and cause of death, and postmortem time interval determinations have also shown huge promise in the field of forensic microbiology (Robinson et al., 2021). However, many challenges still exist in adopting microbial-based body fluid detection. Variation of inter and intra-individual microflora suggests that microbial communities are strongly influenced by many factors, such as geography, time, season, health, diet, genetic factors, anti-inflammatory drugs, and the lifestyle of an individual. Hence, such parameters need to be evaluated before considering a sample for analysis and detection. Another important drawback of the microbial-based approach is its omnipresent nature. Rapid contamination of microorganisms in a biological sample makes the analysis more complicated to distinguish between endogenous and exogenous microflora. Thus, a thorough study on these aspects needs to be carried out before developing a microbial-based technique for human body fluid detection. The detailed pros and cons of available techniques with potential use in body fluid identification have been summarized in Table 22.2.

Table 22.2 A comparative account of the techniques used for the detection of body-fluids with their respective pros and cons.

Sl. no.	Techniques	Body fluid/targets	Advantages	Limitations	References
1	Alternate light source	Blood, semen, vaginal secretions, tears, urine and saliva	Nondestructive Efficient Simple Noninvasive Allows visualization of nonvisible stains or stains on dark surfaces	Limited to use by skilled personnel possessing the knowledge of wavelength selectors Low specificity Poor results when applied on dark fabrics that have been washed Reported to give false-positive result Cannot predict human specificity	Aparna and Iyer (2021), Miranda et al. (2014), Santucci et al. (1999), Vandenberg and van Oorschot, (2006)
	Catalytic tests	Blood, semen, saliva, urine	Simple Fast results Highly sensitive Few tests are nondestructive in nature Inexpensive	False-positive Chemicals like benzidine are known carcinogens Generally presumptive test in nature	An et al. (2012)
	Microscopy	Peripheral blood semen, menstrual blood, vaginal stains, saliva	Easy to use Reliable User friendly	For semen gives false-negative results in natural azoospermic individuals and vasectomized people Can be invasive Sample can't be reused for other tests Time-consuming Using SEM coupled with EDX for preliminary screening of body fluid can be a huge burden of resources on the laboratory	An et al. (2012), Santucci et al. (1999), Scimeca et al. (2018)

Continued

Table 22.2 A comparative account of the techniques used for the detection of body-fluids with their respective pros and cons.—cont'd

Sl. no.	Techniques	Body fluid/targets	Advantages	Limitations	References
	Protein markers	Human peripheral blood, human menstrual blood, semen, vaginal secretions, urine, saliva, sweat	Human specific Diversity of markers Stability over time Tissue-specific posttranslational modifications Can be useful in detecting body fluid in aged forensic samples with high specificity and sensitivity Mixed samples can be analyzed	Cross-reactivity In a mixture of body fluid, detection of the alpha-amylase 1 biomarker in saliva gets interfered by the presence of both blood and semen	de Beijer et al. (2018), Igoh et al. (2015), Lee et al. (2011), Legg et al. (2014), Sakurada et al. (2008)
	mRNA markers	Peripheral blood, menstrual blood, semen, saliva, vaginal secretions	High specificity High sensitivity mRNA markers are stable in nature (remains in dried samples even if samples are aged) Exhibits successful and reliable amplification in much older stains (16-year-old blood stains) Possibility of simultaneous extraction of mRNA and DNA from the same stain as sample is limited Allows detection of several body fluids in one multiplex reaction	Heat and humidity appear to be detrimental to RNA stability False-positive results in saliva sample RNA-based assays still suffer from the fact that RNA is less stable that DNA because of the ubiquitously present RNases	Cossu et al. (2009b), Hanson and Ballantyne (2013), Juusola and Ballantyne (2005), Nussbaumer et al. (2006), Virkler and Lednev (2009), Zubakov et al. (2008)

DNA methylation markers	Blood (venous), saliva, semen, vaginal fluid, urine, menstrual blood	High sensitivity and specificity Requires minute amount of sample Highly applicable and robust assay Analysis of multiple tissues in a single test Numerical, user-independent results DNA-based assay does not necessarily consume physical material (if multiplexed) because it targets extracted DNA Methylation loci and STR typing can be analyzed in a single reaction using commercially available STR profiling kits	Requires skilled expertize Some methylated sequences might not have restriction enzyme sites; methylated sequences that are not present in restriction sites will not be detected Natural variability of methylation levels and stochastic PCR effects need to be considered Genomic DNA can be degraded during bisulfite treatment; a bisulfite-based methylation SNaPshot assay may consume more samples than MSRE-PCR	Frumkin et al. (2011), Gomaa et al. (2017), Lee et al. (2016)
Spectroscopic techniques	Peripheral blood, vaginal fluid, semen, saliva, urine, sweat, menstrual blood	Accurate Sensitive technique Specific Nondestructive Highly reliable Rapid Objective Eco-friendly No or minimal sample preparation required	If the sample is contaminated it cannot be analyzed Water can create interference Liquid samples can't be analyzed due to hydroxyl peaks in the amide regions Many conditions can	Sharma et al. (2020)

Continued

Table 22.2 A comparative account of the techniques used for the detection of body-fluids with their respective pros and cons.—cont'd

Sl. no.	Techniques	Body fluid/targets	Advantages	Limitations	References
				interfere with the spectral results such as water submersion, sunlight exposure, heating, and rust	
	microRNA markers	Nasal secretions, venous blood, saliva, semen, vaginal secretion, urine, menstrual blood	Sensitive and specific approach Stability in samples over long periods Intrinsically small size of miRNAs makes them less prone to degradation by environmental factors	Difficult to discriminate fluids with similar traits such as saliva and vaginal secretion	Fujimoto et al. (2019), Rocchi et al. (2020)
	Circular RNAs	Human seminal fluid, human blood, saliva, vaginal cellular material, menstrual blood	These molecules are closed circular structures so they are very stable Potential to identify fluids in degraded samples	Vaginal cellular material and menstrual blood could not be distinguished	Liu et al. (2019), Memczak et al. (2015), Sijen and Harbison (2021)
	Microbial technique	Vaginal fluid, peripheral blood, saliva, semen, menstrual blood, sweat, urine, feces	Abundance of microflora Stability Signature microflora Commensals in different tissue types	Vaginal and menstrual blood share common microbial signatures Per sample analysis cost is high Variation of interand intraindividual microflora Omnipresent nature of microbes	Castillo et al. (2019), Giampaoli et al. (2014), Neugent et al. (2020), Peterson et al. (2016), Yao et al. (2020)

Rapid techniques (RSID kit, nanotrap sg, SALIgAE, etc)	Human saliva, primate blood, semen, menstrual blood	Fast and efficient The on-site test can be done Performs well on aged and degraded samples	Cross-reactivity Low sensitivity	Laffan et al. (2011), Old et al. (2012), Rahman et al. (2020), Sato et al. (2007), Spalding et al. (2003)
Quantum dot molecular beacon-based techniques	Blood, semen, saliva	Rapid technique Confirmatory test Less time consuming Human-specific Requires less training as compared to PCR-based analysis	Difficulty in multiplexing	Young et al. (2017)
Immunological tests	Human blood, semen, saliva, menstrual blood	Highly sensitive Convenient and rapid Highly specific and easy to use Results are easy to interpret The sample material remains suitable for DNA extraction and profiling	False positives results are common Nonspecificity	An et al. (2012), Holtkötter et al. (2018)
ELISA method	Saliva, vaginal fluid, urine, sweat	Low cost High sensitivity and specificity	Sensitivity and specificity is dependent on the antigen used	Legg et al. (2014), Quarino et al. (2005), Virkler & Lednev (2009)

Conclusions

Body-fluid detection is one of the most important aspects of forensic analysis. Though the primary goal of analyzing any biological sample is individualization, the same sample should undergo its own detection of origin. Hence, the forensic community is in need of a technique that can simultaneously work for body fluid detection as well as individualization. Besides, forensic samples are always precious. Hence, nondestructive tests for body fluid detection are preferred. Biological samples, including body fluids, are mostly found at the crime scene. Before collecting these evidentiary materials, the investigating agency likes to confirm the nature of these body fluids/stains at the crime scene itself. Hence, the forensic science community is also in requirement of portable techniques for body fluid detection at the crime scene. The available techniques for body fluid detection are either not confirmatory or not specific to human origin. Hence, the most suitable technique for body fluid detection should possess characteristic features such as noncross reactivity, human-specificity, portability to the crime scene, and a confirmative nature. With the advancement of technologies, many methodologies have been discovered over the years to detect human-specific body fluids. However, all the techniques described here have their respective pros and cons. Though molecular methods of body fluid detection are still in their developing stages, they have shown huge promise for their applicability in crime scene and laboratory-based investigations of body fluids.

List of abbreviations

ALS Alternate light sources.
ATR-FTIR Attenuated total reflection Fourier transform infrared spectroscopy.
BPPs Bipyrimidine photoproducts.
circRNAs Circular RNAs.
CNV Copy number variation.
DMAC p-dimethylaminocinnamaldehyde.
DNA Deoxyribo nucleic acid.
ELISA Enzyme-linked immunosorbent assay.
ESI-LC-MS Electrospray ionization liquid chromatography-mass spectrometry.
GAPDH Glyceraldehyde-3-phosphate dehydrogenase.
HPLC High-performance liquid chromatography.
LDH Lactate dehydrogenase.
mfDNA Microflora DNA.
NGS Next-generation sequencing.
PAS Periodic acid-Schiff.
PRM1 Protamine-1.
RT-PCR Reverse transcription PCR.
SAP Seminal acid phosphatase.
SEM Scanning Electron Microscopy.
SEMG1 semenogelin-1.
THP Tamm-Horsfall glycoprotein.
VAP Vaginal acid phosphatase.

References

Akutsu, T., Motani, H., Watanabe, K., Iwase, H., & Sakurada, K. (2012). Detection of bacterial 16S ribosomal RNA genes for forensic identification of vaginal fluid. *Legal Medicine, 14*, 160–162. https://doi.org/10.1016/j.legalmed.2012.01.005

Akutsu, T., Watanabe, K., Takayama, A., & Sakurada, K. (2017). Quantitative evaluation of candidate and development of a multiplex RT-PCR assay for the forensic identification of vaginal fluid. *Forensic Science International Genetics, 6*, e211–e213.

Akutsu, T., Yokota, I., Watanabe, K., & Sakurada, K. (2020). Development of a multiplex RT-PCR assay and statistical evaluation of its use in forensic identification of vaginal fluid. *Legal Medicine, 45*, 101715.

Aldunate, M., Tyssen, D., Johnson, A., Zakir, T., Sonza, S., Moench, T., Cone, R., & Tachedjian, G. (2013). Vaginal concentrations of lactic acid potently inactivate HIV. *Journal of Antimicrobial Chemotherapy, 68*, 2015–2025. https://doi.org/10.1093/jac/dkt156

Anderson, D. J., Politch, J. A., Pudney, J., Marquez, C. I., Snead, M. C., & Mauck, C. (2012). A quantitative glycogen assay to verify use of self-administered vaginal swabs. *Sexually Transmitted Diseases, 39*, 949–953. https://doi.org/10.1097/OLQ.0b013e31826e880e

An, J. H., Shin, K. J., Yang, W. I., & Lee, H. Y. (2012). Body fluid identification in forensics. *BMB Reports, 45*, 545–553. https://doi.org/10.5483/bmbrep.2012.45.10.206

Aparna, R., & Iyer, R. S. (2021). Detection and visualization of human tears using alternate light sources for forensic purposes. *Science and Justice, 61*, 771–778. https://doi.org/10.1016/j.scijus.2021.10.001

Banerjee, P. K. (1983). Detection of acid phosphatase in seminal stains by a modified magneson II reagent. *Journal of the Institution of Chemists (India), 55*, 23–25.

Bauer, M., & Patzelt, D. (2003). Protamine mRNA as molecular marker for spermatozoa in semen stains. *International Journal of Legal Medicine, 117*, 175–179.

de Beijer, R. P., de Graaf, C., van Weert, A., van Leeuwen, T. G., Aalders, M. C. G., & van Dam, A. (2018). Identification and detection of protein markers to differentiate between forensically relevant body fluids. *Forensic Science International, 290*, 196–206. https://doi.org/10.1016/j.forsciint.2018.07.013

Bremmer, R. H., Edelman, G., Vegter, T. D., Bijvoets, T., & Aalders, M. C. G. (2011). Remote spectroscopic identification of bloodstains. *Journal of Forensic Sciences, 56*, 1471–1475. https://doi.org/10.1111/j.1556-4029.2011.01868.x

Castella, V., Dimo-Simonin, N., Brandt-Casadevall, C., Robinson, N., Saugy, M., Taroni, F., & Mangin, P. (2006). Forensic identification of urine samples: A comparison between nuclear and mitochondrial DNA markers. *International Journal of Legal Medicine, 120*, 67–72.

Castillo, D. J., Rifkin, R. F., Cowan, D. A., & Potgieter, M. (2019). The healthy human blood microbiome: Fact or fiction? *Frontiers in Cellular and Infection Microbiology*. https://doi.org/10.3389/fcimb.2019.00148

Chatterjee, S. (2019). Saliva as a forensic tool. *Journal of Forensic Dental Sciences, 11*, 1–4. https://doi.org/10.4103/jfo.jfds_69_18

Choi, A., Shin, K. J., & Yang, W. I. (2014). Body fluid identification by integrated analysis of DNA methylation and body fluid-specific microbial DNA. *International Journal of Legal Medicine, 128*, 33–41.

Colotelo, A. H., Cooke, S. J., & Smokorowski, K. E. (2009). Application of forensic techniques to enhance fish conservation and management: Injury detection using presumptive tests for blood. *Endangered Species Research, 9*, 169–178. https://doi.org/10.3354/esr00178

Cossu, C., Germann, U., Kratzer, A., Bär, W., & Haas, C. (2009a). How specific are the vaginal secretion mRNA-markers HBD1 and MUC4. *Forensic Science International Genetics Supplement Series, 2*, 536–537. https://doi.org/10.1016/j.fsigss.2009.08.063

Cossu, C., Germann, U., Kratzer, A., Bär, W., & Haas, C. (2009b). How specific are the vaginal secretion mRNA-markers HBD1 and MUC4? *Forensic Science International Supplement Series, 2*, P536–P537.

Counsil, T. I., & McKillip, J. L. (2010). Forensic blood evidence analysis using RNA targets and novel molecular tools. *Biologia, 65*, 175–182.

Dimitriadis, E., Kalogeropoulos, T., Velaeti, S., Sotiriou, S., Vassiliou, E., Fasoulis, L., Klapsas, V., Synesiou, M., Apostolaki, A., Trangas, T., & Pandis, N. (2013). Study of genetic and epigenetic alterations in urine samples as diagnostic markers for prostate cancer. *Anticancer Research, 33*, 191–197.

Divall, G. B. (1984). A new peptidase isozyme which may assist in the identification of vaginal debris. *Forensic Science International, 24*, 239–246.

Dobay, A., Haas, C., Fucile, G., Downey, N., Morrison, H. G., Kratzer, A., & Arora, N. (2019). Microbiome-based body fluid identification of samples exposed to indoor conditions. *Forensic Science International Genetics, 40*, 105–113. https://doi.org/10.1016/j.fsigen.2019.02.010

Elkins, K. M. (2011). Rapid presumptive "fingerprinting" of body fluids and materials by ATR FT-IR spectroscopy. *Journal of Forensic Sciences, 56*, 1580–1587. https://doi.org/10.1111/j.1556-4029.2011.01870.x

Fiedler, A., Rehdorf, J., Hilbers, F., Johrdan, L., Stribl, C., & Benecke, M. (2008). Detection of semen (human and boar) and saliva on fabrics by a very high-powered UV-/VIS light source. *The Open Forensic Science Journal, 1*, 12–15. https://doi.org/10.2174/1874402800801010012

Fleming, R. I., & Harbison, S. A. (2010). The use of bacteria for the identification of vaginal secretions. *Forensic Science International Genetics, 4*, 311–315. https://doi.org/10.1016/j.fsigen.2009.11.008

Frascione, N., Gooch, J., & Daniel, B. (2013). Enabling fluorescent biosensors for the forensic identification of body fluids. *Analyst, 138*, 7279–7288. https://doi.org/10.1039/C3AN01372C

Frumkin, D., Wasserstrom, A., Budowle, B., & Davidson, A. (2011). DNA methylation-based forensic tissue identification. *Forensic Science International Genetics, 5*, 517–524. https://doi.org/10.1016/j.fsigen.2010.12.001

Fujimoto, S., Manabe, S., Morimoto, C., Ozeki, M., Hamano, Y., Hirai, E., Kotani, H., & Tamaki, K. (2019). Distinct spectrum of microRNA expression in forensically relevant body fluids and probabilistic discriminant approach. *Scientific Reports*. https://doi.org/10.1038/s41598-019-50796-8

Gaensslen, R. E. (1983). *Sourcebook in forensic serology, immunology, and biochemistry*. Washington, DC: U.S. Department of Justice.

Giampaoli, S., Alessandrini, F., Berti, A., Ripani, L., Choi, A., Crab, R., De Vittori, E., Egyed, B., Haas, C., Lee, H. Y., Korabecná, M., Noel, F., Podini, D., Tagliabracci, A., Valentini, A., & Spica, V. R. (2014). Forensic interlaboratory evaluation of the for fluid kit for vaginal fluids identification. *Journal of Forensic Legal Medicine, 21*, 60–63. https://doi.org/10.1016/j.jflm.2013.10.016

Giampaoli, S., Berti, A., Valeriani, F., Gianfranceschi, G., Piccolella, A., Buggiotti, L., Rapone, C., Valentini, A., Ripani, L., & Spica, V. R. (2012). Molecular identification of vaginal fluid by microbial signature. *Forensic Science International Genetics, 6*, 559–564. https://doi.org/10.1016/j.fsigen.2012.01.005

Gomaa, R., Salehi, J., & Behl, S. (2017). DNA methylation as a biomarker for body fluid identification. *Arab Journal of Forensic Science Forensic Medicine, 1*. https://doi.org/10.26735/16586794.2017.001

Gonzalez-Quintela, A., Alende, R., Gude, F., Campos, J., Rey, J., Meijide, L. M., Fernandez-Merino, C., & Vidal, C. (2008). Serum levels of immunoglobulins (IgG, IgA, IgM) in a general adult population and their relationship with alcohol consumption, smoking and common metabolic abnormalities. *Clinical and Experimental Immunology, 151*, 42–50.

Hall, A., Sims, L. M., & Ballantyne, J. (2014). Assessment of DNA damage induced by terrestrial UV irradiation of dried bloodstains: Forensic implications. *Forensic Science International Genetics, 8*, 24–32. https://doi.org/10.1016/j.fsigen.2013.06.010

Hanson, E. K., & Ballantyne, J. (2013). Rapid and inexpensive body fluid identification by RNA profiling-based multiplex high resolution melt (HRM) analysis. *F1000Research, 2*, 281. https://doi.org/10.12688/f1000research.2-281.v2

Hanssen, E. N., Avershina, E., Rudi, K., Gill, P., & Snipen, L. (2017). Body fluid prediction from microbial patterns for forensic application. *Forensic Science International Genetics, 30*, 10–17. https://doi.org/10.1016/j.fsigen.2017.05.009

Harbison, S., & Fleming, R. (2016). Forensic body fluid identification: State of the art. *Research and Reports in Forensic Medical Science, 6*, 11–23.

Hausmann, R., Baltzer, M., & Schellmann, B. (1996). The forensic value of the immunohistochemical detection of oestrogen receptors in vaginal epithelium. *International Journal of Legal Medicine, 109*, 10–13.

Holtkötter, H., Filho, C. R. D., Schwender, K., Stadler, C., Vennemann, M., Pacheco, A. C., & Roca, G. (2018). Forensic differentiation between peripheral and menstrual blood in cases of alleged sexual assault-validating an immunochromatographic multiplex assay for simultaneous detection of human hemoglobin and D-dimer. *International Journal of Legal Medicine, 132*, 683–690. https://doi.org/10.1007/s00414-017-1719-y

Hsu, L., Power, D., & Upritchard, J. (2012). Amplification of oral *Streptococcal* DNA from human incisors and bite marks. *Current Microbiology, 65*, 207–211.

Hurley, I. P., Cook, R., Laughton, C. W., Pickles, N. A., Ireland, H. E., & Williams, J. H. H. (2009). Detection of human blood by immunoassay for applications in forensic analysis. *Forensic Science International, 190*, 91–97.

Igoh, A., Doi, Y., & Sakurada, K. (2015). Identification and evaluation of potential forensic marker proteins in vaginal fluid by liquid chromatography/mass spectrometry. *Analytical and Bioanalytical Chemistry, 407*, 7135–7144.

Jadoon, S., Karim, S., Akram, M. R., Khan, A. K., Zia, M. A., Siddiqi, A. R., & Murtaza, G. (2015). Recent developments in sweat analysis and its applications. *International Journal Analyst Chemistry, 2015*, 164974. https://doi.org/10.1155/2015/164974

Jung, J. Y., Yoon, H. K., An, S., Lee, J. W., Ahn, E. R., Kim, Y. J., Park, H. C., Lee, K., Hwang, J. H., & Lim, S. K. (2018). Rapid oral bacteria detection based on real-time PCR for the forensic identification of saliva. *Scientific Reports, 8*, 10852. https://doi.org/10.1038/s41598-018-29264-2

Juusola, J., & Ballantyne, J. (2005). Multiplex mRNA profiling for the identification of body fluids. *Forensic Science International, 152*, 1–12. https://doi.org/10.1016/j.forsciint.2005.02.020

Kearse, K. P. (2020). Ultraviolet 365 as an alternative light source for detection of blood serum. *Journal of Forensic Sciences, 65*, 1716–1721. https://doi.org/10.1111/1556-4029.14439

Kobus, H., Silenieks, E., & Scharnberg, J. (2002). Improving the effectiveness of fluorescence for the detection of semen stains on fabrics. *Journal of Forensic Sciences, 47*, 819–823. https://doi.org/10.1520/JFS15467J

Kondracki, S., Wysokińska, A., Kania, M., & Górski, K. (2017). Application of two staining methods for sperm morphometric evaluation in domestic pigs. *Journal of Veterinary Research, 61*, 345–349. https://doi.org/10.1515/jvetres-2017-0045

Kosa, F., Rengei, B., & Szendrenyi, J. (1987). New criminalistic important method: Study of cystine aminopeptidase activity for determination of sperm stain. *Krim Forensic Wiss, 67–68*, 123–125.

Laffan, A., Sawyer, I., Quinones, I., & Daniel, B. (2011). Evaluation of semen presumptive tests for use at crime scenes. *Medicine, Science & the Law*. https://doi.org/10.1258/msl.2010.010040

Leake, S. L., Pagni, M., Falquet, L., Taroni, F., & Greub, G. (2016). The salivary microbiome for differentiating individuals: Proof of principle. *Microbes and Infection, 18*, 399–405. https://doi.org/10.1016/j.micinf.2016.03.011

Lee, H. Y., Jung, S. E., Lee, E. H., Yang, W. I., & Shin, K. J. (2016). DNA methylation profiling for a confirmatory test for blood, saliva, semen, vaginal fluid and menstrual blood. *Forensic Science International Genetics, 24*, 75–82. https://doi.org/10.1016/j.fsigen.2016.06.007

Lee, H. Y., Park, M. J., Choi, A., An, J. H., Yang, W. I., & Shin, K. J. (2012). Potential forensic application of DNA methylation profiling to body fluid identification. *International Journal of Legal Medicine, 126*, 55–62. https://doi.org/10.1007/s00414-011-0569-2

Lee, Y. H., Zhou, H., Reiss, J. K., Yan, X., Zhang, L., Chia, D., & Wong, D. T. W. (2011). Direct saliva transcriptome analysis. *Clinical Chemistry, 57*, 1295–1302. https://doi.org/10.1373/clinchem.2010.159210

Legg, K. M., Powell, R., Reisdorph, N., Reisdorph, R., & Danielson, P. B. (2014). Discovery of highly specific protein markers for the identification of biological stains. *Electrophoresis, 35*, 3069–3078. https://doi.org/10.1002/elps.201400125

Li, R. (2008). *Forensic biology*. Boca Raton: CRC Press.

Liu, B., Song, F., Yang, Q., Zhou, Y., Shao, C., Shen, Y., Zhao, Z., Tang, Q., Hou, Y., & Xie, J. (2019). Characterization of tissue-specific biomarkers with the expression of circRNAs in forensically relevant body fluids. *International Journal of Legal Medicine, 133*, 1321–1331. https://doi.org/10.1007/s00414-019-02027-y

López, C. D., González, D. M., Haas, C., Vidaki, A., & Kayser, M. (2020). Microbiome-based body site of origin classification of forensically relevant blood traces. *Forensic Science International Genetics, 47*, 102280. https://doi.org/10.1016/j.fsigen.2020.102280

Matuszewski, M., Szymańska, B., Długosz, A., Małkiewicz, B., Dembowski, J., & Piwowar, A. (2018). Preliminary evaluation of the diagnostic usefulness of uroplakin 2 with an assessment of the antioxidant

potential of patients with bladder cancer. *BioMed Research International, 2018*, 8693297. https://doi.org/10.1155/2018/8693297

McFarland, L. V. (2000). Normal flora: Diversity and functions. *Microbial Ecology in Health and Disease, 12*, 193–207. https://doi.org/10.1080/08910600050216183

Memczak, S., Papavasileiou, P., Peters, O., & Rajewsky, N. (2015). Identification and characterization of circular RNAs as a new class of putative biomarkers in human blood. *PLoS One, 10*, e0141214. https://doi.org/10.1371/journal.pone.0141214

Miranda, G. E., Prado, F. B., Delwing, F., & Daruge, E., Jr. (2014). Analysis of the fluorescence of body fluids on different surfaces and times. *Science & Justice, 54*, 427–431. https://doi.org/10.1016/j.scijus.2014.10.002

Mishra, A., Gondhali, U., & Choudhary, S. (2021). Potential of DNA technique-based body fluid identification. In H. R. Dash, P. Shrivastava, & J. A. Lorente (Eds.), *Handbook of DNA profiling*. Springer.

Nader, L. M., Mohamed, A. S., Ismail, M. M., Fatah, D. S. A., & Ibrahim, T. M. (2015). Molecular biomarkers: The development of MRNA multiplex RT-PCR assay for the definitive identification of semen. *MOJ Toxicology, 1*, 169–177. https://doi.org/10.15406/mojt.2015.01.00025

Neugent, M. L., Hulyalkar, N. V., Nguyen, V. H., Zimmern, P. E., & De Nisco, N. J. (2020). Advances in understanding the human urinary microbiome and its potential role in urinary tract infection. *mBio, 11*, e00218–e00220. https://doi.org/10.1128/mBio.00218-20

Nussbaumer, C., Gharehbaghi-Schnell, E., & Korschineck, I. (2006). Messenger RNA profiling: A novel method for body fluid identification by real-time PCR. *Forensic Science International, 157*, 181–186. https://doi.org/10.1016/j.forsciint.2005.10.009

Old, J., Schweers, B. A., Boonlayangoor, P. W., Fischer, B., Miller, K. W. P., & Reich, K. (2012). Developmental validation of RSID semen: A lateral flow immunochromatographic strip test for the forensic detection of human semen. *Journal of Forensic Sciences, 57*, 489–499. https://doi.org/10.1111/j.1556-4029.2011.01968.x

Oliveira, M., & Amorim, A. (2018). Microbial forensics: New breakthroughs and future prospects. *Applied Microbiology and Biotechnology, 102*, 10377–10391. https://doi.org/10.1007/s00253-018-9414-6

Ong, S. Y., Wain, A., Groombridge, L., & Grimes, E. (2012). Forensic identification of urine using the DMAC test: A method validation study. *Science & Justice, 52*, 90–95. https://doi.org/10.1016/j.scijus.2011.07.007

Pang, B. C. M., & Cheung, B. K. K. (2007). Identification of human semenogelin in membrane strip test as an alternative method for the detection of semen. *Forensic Science International, 169*, 27–31.

Park, J. L., Kwon, O. H., Kim, J. H., Yoo, H. S., Lee, H. C., Woo, K. M., Kim, S. Y., Lee, S. H., & Kim, Y. S. (2014). Identification of body fluid-specific DNA methylation markers for use in forensic science. *Forensic Science International Genetics, 13*, 147–153. https://doi.org/10.1016/j.fsigen.2014.07.011

Passi, N., Garg, R. K., Yadav, M., Singh, R. S., & Kharoshah, M. A. (2012). Effect of luminol and bleaching agent on the serological and DNA analysis from bloodstain. *Egypt Journal of Forensic Science, 2*, 54–61. https://doi.org/10.1016/j.ejfs.2012.04.003

Peterson, R. A., Gueniche, A., de Beaumais, S. A., Breton, L., Csiba, M. D., & Packer, N. H. (2016). Sweating the small stuff: Glycoproteins in human sweat and their unexplored potential for microbial adhesion. *Glycobiology, 26*, 218–229. https://doi.org/10.1093/glycob/cwv102

Quaak, F. C. A., de Graaf, M. L. M., Weterings, R., & Kuiper, I. (2017). Microbial population analysis improves the evidential value of faecal traces in forensic investigations. *International Journal of Legal Medicine, 131*, 45–51. https://doi.org/10.1007/s00414-016-1390-8

Quarino, L., Dang, Q., Hartmann, J., & Moynihan, N. (2005). An ELISA method for the identification of salivary amylase. *Journal of Forensic Sciences, 50*, 873–876. PMID: 16078490.

Quickenden, T. I., & Creamer, J. I. (2001). A study of common interferences with the forensic luminol test for blood. *Luminescence, 16*, 295–298. https://doi.org/10.1002/bio.657

Rahman, A., Khater, S., & Attia, E. (2020). Rapid stain identification (RSID-TMSemen): A rapid tool for seminal fluid detection. *Ain Shams Journal Forensic Mediine Clinical Toxicology, 35*, 73–80. https://doi.org/10.21608/ajfm.2020.111862

Ranjan, R., Rani, A., Metwally, A., McGee, H. S., & Perkins, D. L. (2016). Analysis of the microbiome: Advantages of whole genome shotgun versus 16S amplicon sequencing. *Biochemical and Biophysical Research Communications, 469*, 967–977.

Reich, K. (2007). *Forensic test for human saliva* (USA).

Reich, K. (2008). *Forensic test for human blood*. USA: Independent Forensics Inc.

Richard, M. L. L., Harper, K. A., Craig, R. L., Onorato, A. J., Robertson, J. M., & Donfack, J. (2012). Evaluation of mRNA marker specificity for the identification of five human body fluids by capillary electrophoresis. *Forensic Science International Genetics, 6*, 452–460.

Robinson, J. M., Pasternak, Z., Mason, C. E., & Elhaik, E. (2021). Forensic applications of microbiomics: A review. *Frontiers in Microbiology*. https://doi.org/10.3389/fmicb.2020.608101

Rocchi, A., Chiti, E., Maiese, A., Turillazzi, E., & Spinetti, I. (2020). MicroRNAs: An update of applications in forensic science. *Diagnosis (Basel, Switzerland), 11*, 32. https://doi.org/10.3390/diagnostics11010032

Sagawa, K., Kimura, A., Saito, Y., Inoue, H., Yasuda, S., Nosaka, M., & Tsuji, T. (2003). Production and characterization of a monoclonal antibody for sweat-specific protein and its application for sweat identification. *International Journal of Legal Medicine, 117*, 90–95. https://doi.org/10.1007/s00414-002-0341-8

Saibaba, G., Srinivasan, M., Aarthy, A. P., Silambarasan, V., & Archunan, G. (2017). Ultrastructural and physico-chemical characterization of saliva during menstrual cycle in perspective of ovulation in human. *Drug Discover Theraptics, 11*, 91–97. https://doi.org/10.5582/ddt.2017.01008

Sakurada, K., Akutsu, T., Fukushima, H., Watanabe, K., & Yoshino, M. (2010). Detection of dermcidin for sweat identification by real-time RT-PCR and ELISA. *Forensic Science International, 194*, 80–84. https://doi.org/10.1016/j.forsciint.2009.10.015

Sakurada, K., Motani, H., Akutsu, T., Ikegaya, H., & Iwase, H. (2008). Identification of vaginal stains by detection of 17β-estradiol. *Canadian Society of Forensic Science Journal, 41*, 13–19. https://doi.org/10.1080/00085030.2008.10757161

Santucci, K. A., Nelson, D. G., McQuillen, K. K., Duffy, S. J., & Linakis, J. G. (1999). Wood's lamp utility in the identification of semen. *Pediatrics, 104*, 1342–1344. https://doi.org/10.1542/peds.104.6.1342

Sato, I., Barni, F., & Yoshiike, M. (2007). Applicability of Nanotrap Sg as a semen detection kit before male-specific DNA profiling in sexual assaults. *International Journal of Legal Medicine, 121*, 315–319. https://doi.org/10.1007/s00414-006-0084-z

Scimeca, M., Bischetti, S., Lamsira, H. K., Bonfiglio, R., & Bonanno, E. (2018). Energy dispersive X-ray (EDX) microanalysis: A powerful tool in biomedical research and diagnosis. *European Journal of Histochemistry, 62*, 2841. https://doi.org/10.4081/ejh.2018.2841

Seta, S. (1977). Application of scanning electron microscopy and energy dispersive x-ray microanalysis to the criminal identification of body fluid stains. *International Criminal Police Review, 307*, 119–123.

Shantha Raju, T., Briggs, J. B., Borge, S. M., & Jones, A. J. S. (2000). Species-specific variation in glycosylation of IgG: Evidence for the species-specific sialylation and branch-specific galactosylation and importance for engineering recombinant glycoprotein therapeutics. *Glycobiology, 10*, 477–486.

Sharma, S., Chophi, R., & Singh, R. (2020). Forensic discrimination of menstrual blood and peripheral blood using attenuated total reflectance (ATR)-Fourier transform infrared (FT-IR) spectroscopy and chemometrics. *International Journal of Legal Medicine, 134*, 63–77. https://doi.org/10.1007/s00414-019-02134-w

Sijen, T., & Harbison, S. (2021). On the identification of body fluids and tissues: A crucial link in the investigation and solution of crime. *Genes, 12*, 1728. https://doi.org/10.3390/genes12111728

Sikirzhytski, V., Sikirzhytskaya, A., & Lednev, I. K. (2012). Multidimensional Raman spectroscopic signature of sweat and its potential application to forensic body fluid identification. *Analytica Chimica Acta, 718*, 78–83. https://doi.org/10.1016/j.aca.2011.12.059

Silva, D. S., Antunes, J., Balamurugan, K., Duncan, G., Alho, C. S., & McCord, B. (2016). Developmental validation studies of epigenetic DNA methylation markers for the detection of blood, semen and saliva samples. *Forensic Science International Genetics, 23*, 55–63.

Song, F., Luo, H., & Hou, Y. (2015). Developed and evaluated a multiplex mRNA profiling system for body fluid identification in Chinese Han population. *Journal Forensic Legal Medicine, 35*, 73–80. https://doi.org/10.1016/j.jflm.2015.08.006

Spalding, R. P. (2003). Identification and characterization of blood and bloodstains. In S. H. James, & J. J. Nordby (Eds.), *Forensic science: An introduction to scientific and investigative techniques* (pp. 181–201). Boca Raton: CRC Press.

Stefanidou, M., Alevisopoulos, G., & Spiliopoulou, C. (2010). Fundamental issues in forensic semen detection. *West Indian Medical Journal, 59*, 280–283.

Takamura, A., Watanabe, K., Akutsu, T., & Ozawa, T. (2018). Soft and robust identification of body fluid using fourier transform infrared spectroscopy and chemometric strategies for forensic analysis. *Scientific Reports, 8*, 8459. https://doi.org/10.1038/s41598-018-26873-9

Takayama, T. K., McMullen, B. A., Nelson, P. S., Matsumura, M., & Fujikawa, K. (2001). Characterization of hK4 (prostase), a prostate-specific serine protease: Activation of the precursor of prostate specific antigen (pro-PSA) and single-chain urokinase type plasminogen activator and degradation of prostatic acid phosphatase. *Biochemistry, 40*, 15341–15348.

Tannock, G. W. (1999). The normal microflora: An introduction. In G. W. Tannock (Ed.), *Medical importance of the normal microflora*. Boston, MA: Springer. https://doi.org/10.1007/978-1-4757-3021-0_1

Uchigasaki, S., Tie, J., Sobashima, E., & Shimada, N. (2018). Genotyping DNA isolated from UV irradiated human bloodstains using whole genome amplification. *Molecular Biology Reports, 45*, 925–929. https://doi.org/10.1007/s11033-018-4240-6

Usman, M., Naseer, A., Baig, Y., Jamshaid, T., Shahwar, M., & Khurshuid, S. (2019). Forensic toxicological analysis of hair: A review. *Egypt Journal Forensic Science, 9*, 17. https://doi.org/10.1186/s41935-019-0119-5

Vandenberg, N., & van Oorschot, R. A. (2006). The use of Polilight in the detection of seminal fluid, saliva, and bloodstains and comparison with conventional chemical-based screening tests. *Journal of Forensic Sciences, 51*, 361–370. https://doi.org/10.1111/j.1556-4029.2006.00065.x

Vennemann, M., Scott, G., Curran, L., Bittner, F., & Tobe, S. S. (2014). Sensitivity and specificity of presumptive tests for blood, saliva and semen. *Forensic Science, Medicine and Pathology, 10*, 69–75. https://doi.org/10.1007/s12024-013-9515-6

Virkler, K., & Lednev, I. K. (2009). Analysis of body fluids for forensic purposes: From laboratory testing to non-destructive rapid confirmatory identification at a crime scene. *Forensic Science International, 188*, 1–17. https://doi.org/10.1016/j.forsciint.2009.02.013

Virkler, K., & Lednev, I. K. (2010). Forensic body fluid identification: The Raman spectroscopic signature of saliva. *Analyst, 135*, 512–517. https://doi.org/10.1039/B919393F

Wagner, G., & Levin, R. J. (1980). Electrolytes in vaginal fluid during the menstrual cycle of coitally active and inactive women. *Journal of Reproduction and Fertility, 60*, 17–27.

Yang, Y., Xie, B., & Yan, J. (2014). Application of Next-generation Sequencing technology in forensic science. *Genomics, Proteomics & Bioinformatics, 12*, 190–197. https://doi.org/10.1016/j.gpb.2014.09.001

Yao, T., Han, X., Guan, T., Wang, Z., Zhang, S., Liu, C., Liu, C., & Chen, L. (2020). Effect of indoor environmental exposure on seminal microbiota and its application in body fluid identification. *Forensic Science International, 314*, 110417. https://doi.org/10.1016/j.forsciint.2020.110417

Yasuda, T., Takeshita, H., & Kishi, K. (1997). Genetic polymorphisms detectable in human urine: Their application to forensic individualization. *Nihon Hoigaku Zasshi, 51*, 407–416.

Young, S. T., Moore, J. R., & Bishop, C. P. (2017). A rapid, confirmatory test for body fluid identification. *Journal of Forensic Sciences, 63*, 511–516. https://doi.org/10.1111/1556-4029.13544

Zubakov, D., Boersma, A. W. M., Choi, Y., van Kuijk, P. F., Wiemer, E. A. C., & Kayser, M. (2010). MicroRNA markers for forensic body fluid identification obtained from microarray screening and quantitative RT-PCR confirmation. *International Journal of Legal Medicine, 124*, 217–226. https://doi.org/10.1007/s00414-009-0402-3

Zubakov, D., Ciemińska, J. C., Kokmeijer, I., Maciejewska, A., Martínez, P., Pawłowski, R., Haas, C., & Kayser, M. (2018). Introducing novel type of human DNA markers for forensic tissue identification: DNA copy number variation allows the detection of blood and semen. *Forensic Sci Int Genet, 36*, 112–118. https://doi.org/10.1016/j.fsigen.2018.06.021

Zubakov, D., Hanekamp, E., Kokshoorn, M., van Ijcken, W., & Kayser, M. (2008). Stable RNA markers for identification of blood and saliva stains revealed from whole genome expression analysis of time-wise degraded samples. *International Journal of Legal Medicine, 122*, 135–142. https://doi.org/10.1007/s00414-007-0182-6

CHAPTER 23

Application of next-generation sequencing technology and investigative genetic genealogy to solving cold cases

Daria Smirnova and Kelly M. Elkins
Department of Chemistry, Forensic Science Program, Towson University, Towson, MD, United States

Introduction

Investigative Genetic Genealogy (IGG) is the approach of combining genetic analysis with traditional criminal investigations. It is an emerging science that examines single nucleotide polymorphisms (SNPs) in the human genome. The terms next-generation sequencing (NGS) and massively parallel sequencing (MPS) describe the newest technologies that sequence DNA and RNA in a more rapid and cost-effective manner (Børsting and Morling, 2015). These sequencing methods have brought about the use of different types of genetic markers, such as SNPs, that can be used in forensic genetic analysis. The use of SNPs as additional genetic markers has led to larger genetic datasets and grown alongside new stronger computational methods and automation, as well as the compilation of additional and larger DNA databases. Decades of these advancements in DNA science and technology have led to IGG becoming one of the most exciting methods used to identify relatives of criminal suspects to crack cold cases. IGG is mainly applied for this purpose; however, it is also used for other investigative questions. A search result on Google for "IGG" yielded nearly 8,320,000 hits (January 16, 2023), including research studies, articles, news stories, and videos demonstrating its widespread popularity. Genetic genealogy has also been used in disaster victim identification cases and missing person identification. Many of these involve the need for large-scale genetic profile comparisons (Chen et al., 2019).

The IGG process employs searching genetic data from privately owned direct-to-consumer (DTC) companies, which store millions of DNA profiles of customers, and dense SNP data consisting of half a million markers to investigate distant relationships (Machado and Silva, 2022). The degree of relatedness is parameter that can be set by investigators to narrow down or widen the search. DTC genetic databases empower IGG, but not all DTC companies are on board with allowing law enforcement access due to customer privacy and other commercial concerns. This makes some of them unavailable for use in IGG. At present, an imbalance exists between the growing scientific usefulness

Next Generation Sequencing (NGS) Technology in DNA Analysis
ISBN 978-0-323-99144-5, https://doi.org/10.1016/B978-0-323-99144-5.00029-9
© 2024 Elsevier Inc.
All rights reserved.

of IGG in forensic casework by the police and prosecution and the actual number of times it is used.

Nonetheless, IGG with NGS has become a powerful tool used in forensic casework to solve cold cases, especially in more recent years and after its use in solving high-profile cases. The well-known case of the 'Golden State Killer' was solved by a ancestry search in a DTC database owned by GEDMatch after the unidentified killer committed a number of murders and more than 50 sexual assaults (SA) over two decades in the 1970 and 1980s (Wickenheiser, 2019). His DNA profile was found on different items of evidence collected from the crime scene, but investigators could not identify him due to the nonavailability of his DNA in any law enforcement databases. The GEDMatch database, originally launched in 2010, was used in 2018 by a California police officer who ran a genealogical search and identified a third cousin match with the culprit's DNA profile from crime scene evidence. The profile match was found to be that of a third cousin of Joseph James DeAngelo, a breakthrough after 40 years. By tracing through his family tree, DeAngelo was narrowed down as a suspect. His sample was taken as a reference, and when the DNA profile was generated, it matched the evidence sample profiles obtained in past cases. The investigation led to his apprehension, and the case led to the expansion of the use of various ancestry-based commercial databases worldwide. Aside from the DeAngelo case, this study collected available data from 65 public news report articles on 80 additional cold cases recently reported to have been solved through the use of IGG and NGS [5–69]. The goals was to summarize all reported cold cases solved using NGS and IGG. It highlights the importance of IGG's current use in solving cold cases and its potential for solving even greater numbers of them and more of them before they turn cold.

Methods

In order to obtain an overview and assess all reported cold cases solved with NGS or IGG, a list of such cases was compiled from all such reported cold cases solved. Multiple internet search engines were utilized to search for the keywords "cold cases solved with…" followed by the words "NGS", "IGG", "investigative genetic genealogy", "forensic genealogy", and "DNA", all in separate searches. "DNA solve" and related phrases were also searched. The primary sources found were news reports and news articles, with a few other sources such as science and law journals. This revealed 80 additional solved cases, which were then reviewed for additional significant data of interest, which was gathered and compiled out of the news articles and reports. Each news article was used to collect data on both the perpetrators and victims. Data collected on the perpetrators includes their name, reported gender, race, age, the type of crime or crimes they committed, when they were committed, the geographical location where they were committed, the year they were arrested and brought to justice, and crimes solved (Table 23.1). Corresponding data was also collected about the victims (Table 23.2).

Table 23.1 Information on perpetrators of cold cases solved with investigative genetic genealogy.

	Name	Gender	Race	State[a]	Year	Crime type	Date solved
1	Steve Branch	M	C	AK	1996	H, SA	2020
2	Lesa Lopez	F	C	CA	1988	Filicide	2020
3	Alan Edward Dean	M	C	WA	1993	Kidnapping, H	2020
4	Michael Allan Carbo Jr	M	C	MN	1986	H, SA	2020
5	William Baer	M	C	FL	1999	H	2020
6	Robert Lynn Bradley	M	C	FL	1991	H	2020
7	Julius William Hill Jr	M	C	FL	1978	H	2020
8	Daniel Nyqvist	M	C	Sweden	2004	H	2020
9	Daniel Alan Anderson	M	AA	DE	1983	H	2020
10	Phillip Wilson	M	AA	CA	1980	H, SA	2020
11	Terry Rasmussen	M	C	NH	1978	H	2020
12	James Curtis Clanton	M	C	CO	1980	H, SA	2020
13	Robert Dale Edwards	M	C	CA	1973	H	2020
14	Bruce Lindahl	M	C	IL	1976	H	2020
15	Unnamed	M	UNK	NC	1984	H	2018
16	Charles Gary Sullivan	M	C	AZ	1970	H, SA	2019
17.1	Horace Van Vaultz Jr.	M	AA	CA	1986	H, SA	2019
17.2	Horace Van Vaultz Jr.	M	AA	CA	1981	H, SA	2019
18	Linda LaRoche	F	C	WI	1999	H	2019
19	Brook Graham	F	C	SC	1989	H	2019
20	John Arthur Getreu	M	C	CA	1974	H	2019
21	Frank Wypych	M	C	WA	1967	H, SA	2019
22	Jeffrey Lynn Hand	M	UNK	IND	1972	H, SA	2019
23	Richard Knapp	M	C	Vancouver	1994	H	2019
24	Terrence Miller	M	C	WA	1972	SA, H	2019
25.1	Cecil Stan Caldwell	M	C	MT	1973	H	2019
25.2	Cecil Stan Caldwell	M	C	MT	1973	H	2019
26.1	Coley McCraney	M	AA	AL	1999	SA, H	2019
26.2	Coley McCraney	M	AA	AL	1999	SA, H	2019

Continued

Table 23.1 Information on perpetrators of cold cases solved with investigative genetic genealogy.—cont'd

	Name	Gender	Race	State[a]	Year	Crime type	Date solved
27.1	Raymond L. Vannieuwenhoven	M	C	WI	1976	H	2019
27.2	Raymond L. Vannieuwenhoven	M	C	WI	1976	SA, H	2019
28.1	Kenneth Earl Day	M	C	MD	1994	SA, H	2019
28.2	Kenneth Earl Day	M	C	MD	1989	SA	2019
29	Thomas Levis Garner	M	AA	FL	1984	H	2019
30.1	Joseph Holt	M	C	CO	1977	SA, H	2019
30.2	Joseph Holt	M	C	CO	1979	SA, H	2019
31	James Neal	M	C	CA	1973	SA, H	2019
32	Steven Downs	M	C	AK	1993	SA, H	2019
33	Jerry McFadden	M	C	OR	1986	SA, H	2019
34	Zachary Bunny	M	C	OR	2006	H	2019
35	Russell Guerrero	M	C	AZ	1990	H	2019
36	Raymond Charles Rowe	M	C	PA	1992	SA, H	2019
37	John D.Miller	M	C	IN	1988	SA, H	2018
38	John A. Getreu	M	C	CA	1973	H	2018
39	David Mabrito	M	C	CA	2007	SA, H	2018
40	Benjamin L.Holmes	M	AA	FL	2001	H	2018
41	Fred Lee Frampton	M	C	MD	2010	H	2019
42	Jerry Lee	M	AA	GA	1997	H	2018
43	Edward Keith Renegal	M	C	KS	1990	SA, H	2018
44	Michael W.Devaughn	M	C	MS	1990	SA, H	2018
45	Robert Brashers	M	C	SC	1990	SA, H	2018
46	Luke Fleming	M	C	FL	1999	SA, H	2020
47	Michael Henslick	M	C	IL	2009	H	2018
48	Mathew Dusseault	M	C	MA	2016	H	2018
49	Tyler Grenon	M	C	MA	2016	H	2018
50	James Otto Earhart	M	C	TX	1981	H	2018
51	Gary Hartman	M	C	WA	1986	SA, H	2018
52.1	William Talbot	M	C	WA	1987	H	2018
52.2	William Talbot	M	C	WA	1987	H	2018

53	Gary Young	M	C	AZ	1981	SA	2020
54	Leroy Jemol Smith	M	AA	OK	2005	SA	2020
55	Mark Dauglas Burns	M	C	UT	1992	SA	2020
56	Gregory Paul Vein	M	C	CA	1997	SA	2019
57	Gilles Warrick	M	AA	SC	1990	SA, H	2019
58	Johnnie B.Greens	M	AA	NC	2009	SA	2019
59	Roy Charles Waller	M	C	NC	1991–2006	SA	2019
60	William Louis Nicholas	M	C	FL	1983	SA	2019
61.1	Marlon Michael Alexander	M	AA	MD	2010	SA	2019
61.2	Marlon Michael Alexander	M	AA	MD	2010	SA	2019
61.3	Marlon Michael Alexander	M	AA	MD	2011	SA	2019
61.4	Marlon Michael Alexander	M	AA	MD	2007	SA	2019
62	Darold Wayne Bowden	M	C	NC	2006	SA	2019
63	Spencer Glen Monnett	M	C	UT	2018	SA	2018

AA, African American; *C*, Caucasian; *H*, homicide; *SA*, sexual assault; *UNK*, Unknown.
[a]For cases outside of the USA, city or country is reported.

Table 23.2 Information on victims of cold cases solved with investigative genetic genealogy.

	Name	Gender	Race	State[a]	Age	Year	Type of assault	Date solved
1	Jessica Baggen	F	C	AK	17	1996	H, SA	2020
2	Baby John Doe	M	C	CA	0	1988	H	2020
3	Melissa Lee	F	C	WA	15	1993	H	2020
4	Nancy Daugherty	F	C	MN	38	1986	H, SA	2020
5	Saad Kawaf	M	C	FL	est.45	1999	H	2020
6	Wanda Deann Kirkum	F	C	FL	18	1991	H	2020
7	Carolyn Cox Rose	F	C	FL	47	1978	H	2020
8	Anna–Lena Svensson	F	C	Sweden	56	2004	H	2020
9	John Muncy	M	C	DE	15	1983	H	2020
10	Robin Brooks	F	C	CA	20	1980	SA, HH	2020
11	Marlyse Honeychurch	F	C	NH	24	1978	H	2020
12	Helene Pruszynski	F	C	CO	21	1980	H, SA	2020
13	Naomi Sanders	F	C	CA	57	1973	H	2020
14	Pamela Maurer	F	C	IL	16	1976	H	2020
15	Reesa Trexler	F	C	NC	15	1984	H, SA	2018
16	Julia Woodward	F	C	NV	21	1979	H, SA	2019
17	Mary Duggan	F	C	CA	22	1986	H, SA	2019
18	Selena Keough	F	C	Montclair	20	1981	H, SA	2019
19	Peggy Lynn Johnson	F	C	IL	18	1999	H	2019
20	Baby Boy	M	UNK	SC	Newborn	1989	H	2019
21	Janet Taylor	F	C	CA	21	1974	H	2019
22	Susan Galvin	F	C	WA	20	1967	H	2019
23	Pam Milam	F	UNK	IN	18–22	1972	H, SA	2019
24	Audrey Hoellein	F	C	OR	26	1994	H	2019
25	Jody Loomis	F	C	WA	20	1972	SA, H	2019
26	Clifford Bernhardt	M	C	MT	20	1973	H	2019
27	Linda Bernhardt	F	C	MT	20	1973	H	2019
28	Tracie Hawlette	F	C	AL	17	1999	SA, H	2019
29	J.B.Beasley	F	C	AL	17	1999	SA, H	2019
30	David Schuldes	M	C	WI	25	1976	H	2019

31	Ellen Matheys	F	C	WI	24	1976	SA, H	2019
32	Le Bich-Thuy,	F	A	MD	42	1994	SA, H	2019
33	Unknown	F	UNK	MD	52	1989	SA	2019
34	Pamela Cahanes	F	C	FL	25	1984	H	2019
35	Bynn Riney	F	C	CO	27	1977	SA, H	2019
36	Carol Anderson	F	C	CO	16	1979	SA, H	2019
37	Linda O'Keefe	F	C	CA	11	1973	SA, H	2019
38	Sophie Sergie	F	A	AK	20	1993	SA, H	2019
39	Anna Marie Hlavka	F	C	OR	18	1986	SA, H	2019
40	Scott Martinez	M	H	OR	47	2006	H	2019
41	Jack Upton	M	C	AZ	30	1990	H	2019
42	Christy Mirack	F	C	PA	25	1992	SA, H	2019
43	April Tinsley	F	C	IN	8	1988	SA, H	2018
44	Leslie Perlov	F	C	CA	21	1973	H	2018
45	Jodine Serrin	F	C	CA	39	2007	SA, H	2018
46	Christine Franke	F	C	FL	25	2001	H	2018
47	Michael A. Temple	M	C	MD	24	2010	H	2018
48	Ann Smith	F	C	GA	28	1997	H	2018
49	Pamela Faye Felkins	F	C	KS	32	1990	SA, H	2018
50	Betty Jones	F	C	MS	65	1990	SA, H	2018
51	Jenny Zitricki	F	C	SC	28	1990	SA, H	2018
52	Deborah Dalzell	F	C	FL	47	1999	SA, H	2020
53	Holly Cassano	F	C	IL	22	2009	H	2018
54	Constance Gauthier	F	C	MA	81	2016	H	2018
55	Virginia Feeman	F	C	TX	40	1981	SA, H	2019
56	Michella Welch	F	C	WA	12	1986	SA, H	2018
57	Tanya Van Cuyleborg	F	C	WA	18	1987	H	2018
58	Jay Cook	M	C	WA	20	1987	H	2018
59	Unknown	F	UNK	AZ	NA	1991	SA	2020
60	Unknown	F	UNK	OK	NA	1993–95	SA	2020
61	Unknown	F	UNK	CA	NA	1997	SA	2019

Continued

Table 23.2 Information on victims of cold cases solved with investigative genetic genealogy.—cont'd

	Name	Gender	Race	State[a]	Age	Year	Type of assault	Date solved
62	Christine Mirzayan	F	C	SC	29	1998	SA, H	2019
63	Unknown	F	UNK	NC	NA	2009	SA	2019
64	Unknown	F	UNK	NC	NA	2006	SA	2019
65	Unknown	F	UNK	FL	12	1983	SA	2019
66	Unknown	F	UNK	MD	25	2007	SA	2019
67	Unknown	F	UNK	MD	68	2010	SA	2019
68	Unknown	F	UNK	MD	86	2010	SA	2019
69	Unknown	F	UNK	MD	68	2011	SA	2019
70	Unknown	F	UNK	MD	NA	2007	SA	2019
71–76	Unknown X6	F	UNK	NC	NA	2006–08	SA	2019
77	Unknown	F	UNK	UT	79	2018	SA	2018
78	Lisa Roberts	F	C	WA	17	1977	H	2020
79	Paul Fronczak	M	C	IL	NA	1964	Kidnapping	2019
80	Horace Jack	M	UNK	SW	49	1976	Unknown	2019

H, homicide; *SA*, sexual assault; *AA*, African American; *C*, Caucasian; *UNK*, Unknown.
[a]For cases outside of the USA, city or country is reported.

Multiple crimes committed by the same individual were denoted with the same number with a dash for each serial numerical count. Because many of the perpetrators committed multiple crimes on multiple victims, the number of perpetrators is less than the number of victims listed, with 63 perpetrators caught and 80 victims' cases solved.

A given news article's publication date was used as an estimate of the date on which the crimes were solved, which is one potential limitation of the accuracy of that value in this research. Likewise, all information and any results interpreted from the sources in this study rely on and are limited by the accuracy of the news articles reporting on the solved cold cases. The information was analyzed using pie charts and bar graphs to perform comparisons based on the categories of data collected. Comparisons are shown for the percentage of crimes committed by the reported gender of the perpetrator (Fig. 23.1), race of the perpetrator (Fig. 23.2), geographical location (Fig. 23.3), gender of victims (Fig. 23.4), and length of time in years that the crime went unsolved (Fig. 23.5). Repeated offenses by the same individual were counted separately in statistics totaling the number of crimes. Not all of the victims and perpetrators were accounted for in each of the statistics because, in some cases, the name, gender, or other details of the crimes remain unknown. For determining any percentiles or other comparisons in any cases with these unknowns, the unknowns were omitted from the count and not included in the calculation. Crimes by category were tabulated in Fig. 23.6.

Results

From the gathered news information, the oldest cold case solved was the kidnapping of Paul Fronczak, which went unsolved for over 55 years. The data is shown in Figs.

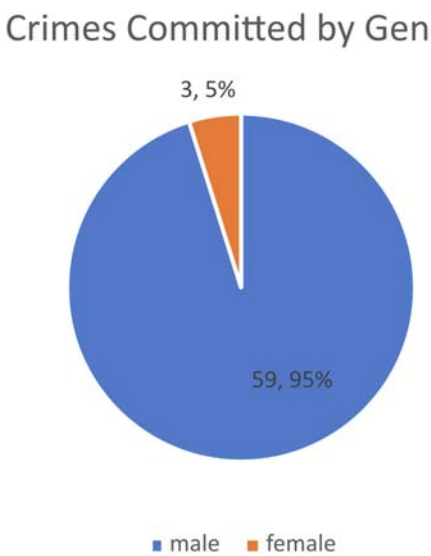


Figure 23.1 Gender comparisons by number and percentage for perpetrators of crimes reported solved.

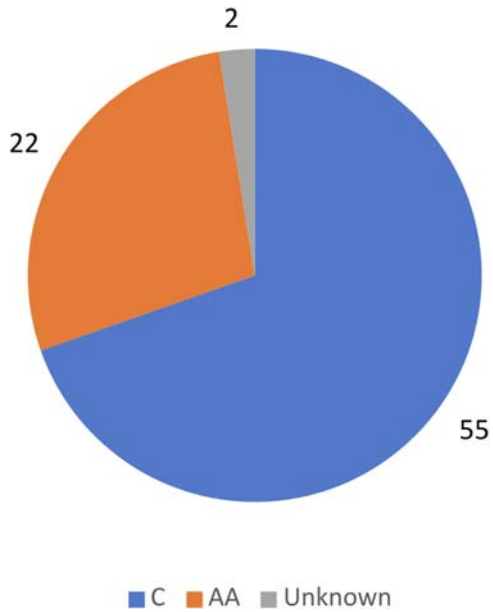


Figure 23.2 Comparison of perpetrators race by number of crimes reported solved.

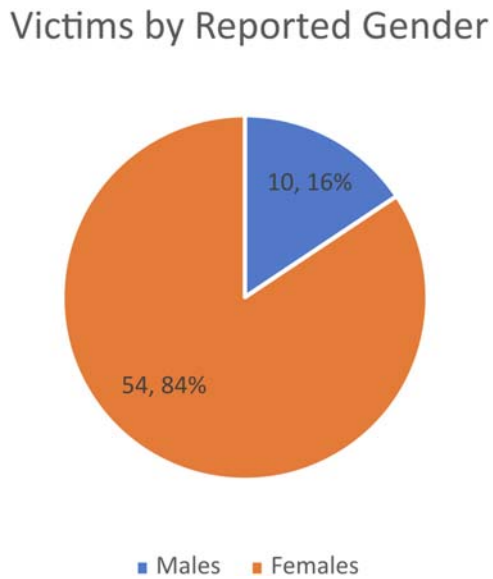


Figure 23.3 Gender of victims by number count and percentage.

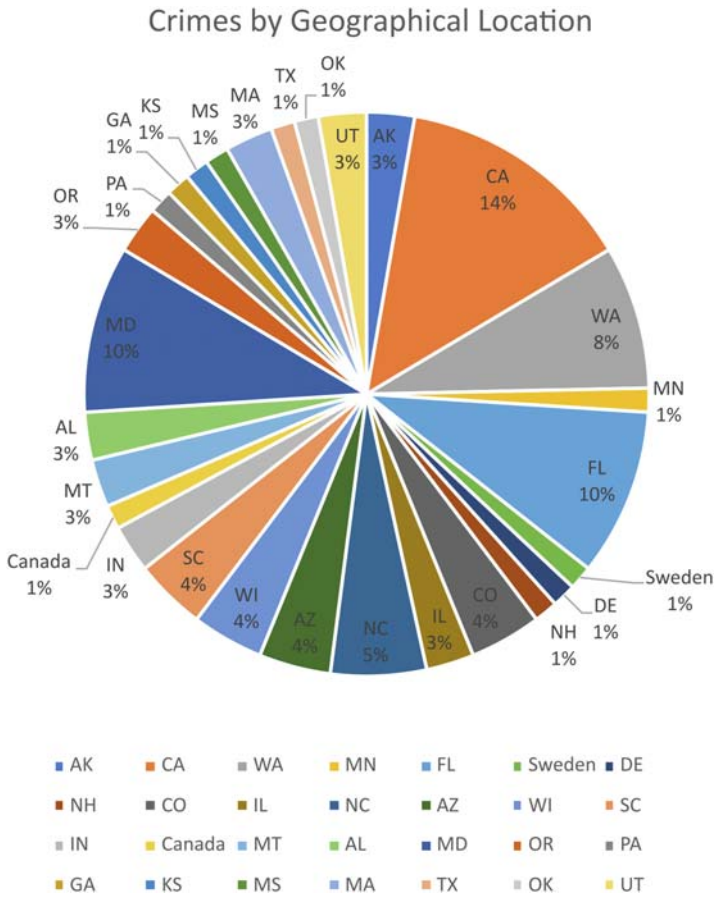


Figure 23.4 Geographical location of crimes by US state/province and country.

23.1–23.6. Other than the single kidnapping case, the cold cases were all sexual assault (SA) and homicide (H) cases, where 95% (n = 59) of cases the crimes were committed by males, specifically Caucasian males 75% (n = 55), African American males 22% (n = 16), and 3% of unknown ancestry (n = 2) (Table 23.1 and Figs. 23.1 and 23.2). Out of the total number of assailants (n = 63), five white males and two African American males were repeat offenders involved in sexual assault cases, homicides, or both. The remaining 5% (n = 2) of cold cases were committed by Caucasian female assailants, which were strictly linked to homicides. The geographic locations of the crimes are shown in Fig. 23.4. The higher case distribution in certain states, such as Maryland was affected by

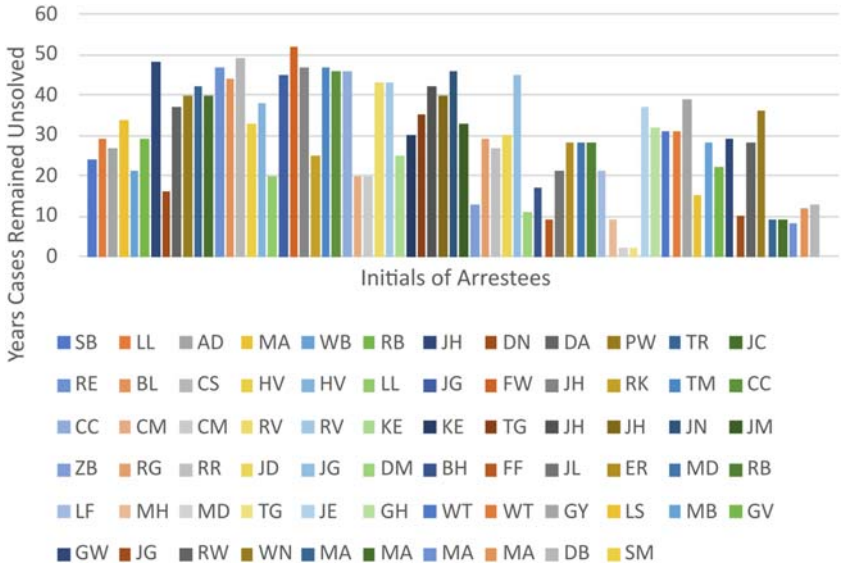

60
50
40
30
20
10
0
Years Cases Remained Unsolved
Initials of Arrestees
SB LL AD MA WB RB JH DN DA PW TR JC
RE BL CS HV HV LL JG FW JH RK TM CC
CC CM CM RV RV KE KE TG JH JH JN JM
ZB RG RR JD JG DM BH FF JL ER MD RB
LF MH MD TG JE GH WT WT GY LS MB GV
GW JG RW WN MA MA MA MA DB SM

Figure 23.5 Comparison of lengths of time crimes remained unsolved.

the serial offenses committed by a single individual. More cold cases were solved overall in large states including California and Florida. The victims (n = 80, ages ranging from age 0 to 86, with an average age of 29.4 and standard deviation of 18.61 years) (Table 23.2 and Fig. 23.3) were gathered using only available news information. Females comprised 66 victims and accounted for 85% of the victims, with 95% of the victims being Caucasian females, while there were also two Asian females, and one Hispanic female victim. In some cases, variables of interest were unavailable and, as such, were not included in analysis. For example, due to privacy concerns the surviving victim's name, age, and race were often undisclosed. However, the living victims' genders were counted and applied in the comparisons, as well as the category of offense. Fig. 23.5 shows the time it took to solve a case. It took 29.1 years on average to link a perpetrator to a crime, with a standard deviation of 13.36 years . It is noted that all cold cases in this study were solved in the years 2018, 2019, and 2020 and do not account for cases that might have been solved more recently. The crime categories were subdivided into homicides (H) (n = 30, 40%), sexual assault (SA) (n = 13, 17%), homicide with sexual assault (H with SA) (n = 31, 41%), kidnapping (n = 1, 1.1%), and unknown categories (n = 1, 1.1%) (Fig. 23.6).

Discussion and conclusion

Investigative genetic genealogy has become an important tool for unlocking cold cases. It is estimated that GEDmatch data contains over 1.2 million individual profiles, with a 90%

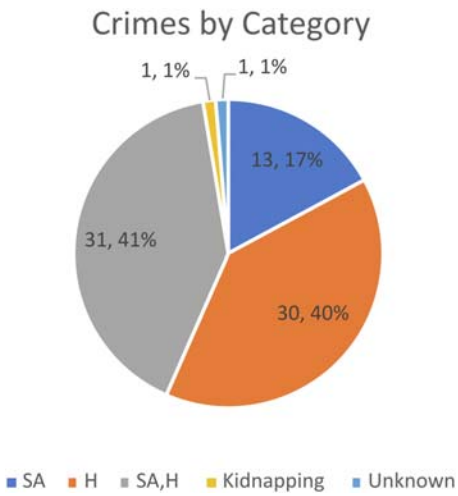


Figure 23.6 Comparison of type of crimes by number and percentages.

chance to identify at least one-third of the cousins of anyone in the population (Khan and Mittelman, 2018). Each cold case described in the articles was based on DNA profiles that were closely matched to the crime scene or evidence through genetic association. The generated data demonstrates that crimes solved by these methods are heavily skewed toward white male perpetrators and young white female victims. The majority of crimes were sexually motivated homicides, which gives some explanation for the majority of victims being dominantly female. The fact that white males are the dominant perpetrators in these solved crimes is possibly explained by the fact that DTC customer databases have primarily Caucasians as customers, making the majority of genetic data received and held by DTCs from Caucasians. This logically leads to catching more perpetrators that are family members of Caucasians. Due to a lack of sufficient evidence, it remained unclear in several repeat offense cases whether all their victims were sexually assaulted before being murdered. Serial killing offenses were also committed mainly by white males. One black male sexually assaulted six women in Maryland but did not murder them. The geographic distribution of crimes was disproportionally affected by the serial offense cases, as opposed to being based on population, socioeconomic, or other factors related to location. All listed cases were solved in the years 2018, 2019, and 2020, with the oldest case remaining unsolved for 55 years and the newest solved in the same year of the offense. The trend of old cases remaining unsolved has significantly decreased in recent years. This is likely, at least in part, due to the growth and increased scientific and technological ability to solve crimes with these databases, enhanced by the ongoing growth and developments in NGS, including the ability to use an ever-growing number of different types of genetic markers such as SNPs and InDels. The trend is also likely due in part to the genetic database GEDmatch being purchased by Verogen Company

in 2019, which then dedicated the platform to forensic investigatory uses (Erlick et al., 2018). Recently, Verogen has ensured that anyone contributing to GEDmatch is explicitly warned that criminal investigators or other people just interested in genealogies will be able to perform DNA comparisons against their data (Greytak et al., 2019). The tool now opens the database to law enforcement if the customers receive this warning and still choose to use the service and contribute their genetic data for this purpose. Additionally, privacy concerns are reduced by the fact that no one is legally required to contribute to a genetic genealogy database. It is purely voluntary, and all the DNA data samples are not in the possession of any government agencies. The majority of success in catching criminals with IGGs on DTC databases has been in cold cases. However, genetic genealogy has just as much ability to generate leads in active cases. In 2018, it was used to identify and arrest a perpetrator in an active sexual assault case that occurred only 3 months prior to the arrest. The individual arrested pled guilty in court in 2019. In January 2023, it was utilized to lead to the arrest in Pennsylvania of a suspect in the murder of four Idaho college students (Sun et al., 2022). While this suspect has not yet been brought to court and found guilty, it shows that investigators can now have access to modern forensic DNA technologies that can quickly generate significant new leads rather than waiting until all other means of investigation are exhausted. IGG with NGS can prevent cases from ever going cold as well as continue to solve increasing numbers of unsolved cold cases. To help with this, opening DTCs with large databases to police access, despite privacy and other concerns will aid in solving morecases. Police and law enforcement, along with academies, universities, and private DTC companies working to develop their relationships with each other and to form standard procedures and protocols, will facilitate more familiar interactions with one another and enable law enforcement greater and quicker access for searches. Police searches of these DNA databases have strongly shown their potential to solve some of the most serious crimes in a quick time. Additional consumer contributions to the databases from all races and ethnic groups make facilitate cases being solved for more victims.

References

Børsting, C., & Morling, N. (2015). Next generation sequencing and its applications in forensic genetics. *Forensic Science International Genetics, 18*, 78–89. https://doi.org/10.1016/j.fsigen.2015.02.002

Chen, P., Zhu, W., Tong, F., Pu, Y., Yu, Y., Huang, S., et al. (2019). Identifying novel microhaplotypes for ancestry inference. *International Journal of Legal Medicine, 133*(4), 983–988. https://doi.org/10.1007/s00414-018-1881-x

Erlick, Y., Shor, T., Pe'er, I., & Carmi, S. (2018). Identity inference of genomic data using long- range familial searches. *Science, 362*(6415), 690–694. https://doi.org/10.1126/science.aau483

Greytak, E. M., Moore, C., & Armentrout, S. L. (2019). Genetic genealogy for cold case and active investigations. *Forensic Science International, 299*, 103–113. https://doi.org/10.1016/j.forsciint.2019.03.039

Khan, R., & Mittelman, D. (2018). Consumer genomics will change your life, whether you get tested or not. *Genome Biology, 19*(120), 1–4. https://doi.org/10.1186/s13059-018-1506-1

Machado, H., & Silva, S. (2022). Investigative Genetic Genealogy: an ethical and privacy assessment framework tool is needed. *Forensic Science Review Genetics, 18*, 17−19.
Sun, R., Bogel-burroughs, N., & Kovaleski, S. F. (2022). Criminology student is charged in 4 university of idaho killings. *The New York Times*. [cited 2023 Jan 29];Available from https://www.nytimes.com/2022/12/30/us/idaho-murders-suspect-bryan-kohberger.html. [cited 2023 Jan 29];Available from.
Wickenheiser, R. (2019). Forensic genealogy, bioethics and the golden state killer case. *Forensic Science International, 1*, 114−125. https://doi.org/10.1016/j.fsisyn.2019.07.003

Case Reports

ABC News. ABC news. [cited 2023 Jan 29];Available from https://abcnews.go.com/US/man-arrested-2009-murder-22-year-mother-type/story?id=57480270.
Amanda Becker FOX11 N. (2019). *Arrest made in 43-year-old marinette county murder case*. [cited 2023 Jan 29];Available from https://fox11online.com/news/local/arrest-made-in-1976-double-murder.
Arrest made in 1986 Chisholm cold case murder of Nancy Daugherty. *Fox21Online*, (2020).
Belt, D. (2018). *Murder charge filed in odenton cold case, 2nd suspect dead*. Odenton-Severn, MD Patch. [cited 2023 Jan 29];Available from https://patch.com/maryland/odenton/murder-charge-filed-odenton-cold-case-2nd-suspect-dead.
Blowers, C. (2018). Police arrest hurricane man for alleged rape of 79-year-old woman in St. George. *St George News*. [cited 2023 Jan 29];Available from https://archives.stgeorgeutah.com/news/archive/2018/07/28/cgb-police-arrest-hurricane-man-for-alleged-rape-of-79-year-old-woman-in-st-george#.Y9cOgXZOm5c. [cited 2023 Jan 29];Available from .
Boots, M. T. (2020). Suspect in 1996 cold-case murder of sitka girl killed himself last week in Arkansas, Troopers Say. *Anchorage Daily News*. [cited 2023 Jan 28];Available from https://www.adn.com/alaska-news/crime-courts/2020/08/11/troopers-announce-theyve-identified-killer-in-1996-sitka-cold-case-murder-of-teenage-girl-arkansas-suspect-kills-himself/.
Bureau FCPDPA. (2022). Cold case victim identified through DNA testing. *Fairfax County Police Department News*. [cited 2023 Jan 29];Available from https://fcpdnews.wordpress.com/2022/12/02/cold-case-victim-identified-through-dna-testing/.
Callaghan, L. (2020). Genealogist uses ancestry website to track down knife killer. *World | The Sunday Times*. [cited 2023 Jan 29];Available from https://www.thetimes.co.uk/article/genealogist-uses-ancestry-website-to-track-down-knife-killer-m60rs0j2l.
Capelouto, Z. H. (2018). *21 years later, police make arrest in shooting death of fulton county woman*. ajc. [cited 2023 Jan 29];Available from https://www.ajc.com/news/years-later-police-make-arrest-shooting-death-fulton-county-woman/BDCqr2HASqidoaUhURqV1J/.
Coello, S. (2019). Bryan murderer exhumed to confirm DNA evidence he was responsible for another decades-old cold case. *Dallas News*. [cited 2023 Jan 29];Available from https://www.dallasnews.com/news/crime/2018/06/27/bryan-murderer-exhumed-to-confirm-dna-evidence-he-was-responsible-for-another-decades-old-cold-case/. [cited 2023 Jan 29];Available from.
Daniel, J., Gross, G. C., & ELF. (2018). Greenville cold case solved: Jenny Zitricki killed by serial killer Robert Brashers in 1990, police say. *The Greenville News*. [cited 2023 Jan 29];Available from https://www.greenvilleonline.com/story/news/crime/2018/10/05/greenville-cold-case-zitricki-murder-first-series-murders-across-three-states/1532153002/.
Desk HTVN. (2019). *Truck-driving preacher charged with killing 2 teenage girls nearly 20 years ago*. New Haven Register. [cited 2023 Jan 29];Available from https://www.nhregister.com/news/article/Police-charge-truck-driving-preacher-with-killing-13699145.php.
DNA leads to arrest of man in 1987 killing of couple in Washington state. CBS News. 2018 [cited 2023 Jan 29];Available from: https://www.cbsnews.com/news/william-earl-talbott-arrested-in-1987-killing-of-tanya-van-cuylenborg-jay-cook-in-washington/.
DNA on plastic spoon leads to Livermore Man's arrest for 1997 rapes. *CBS News*, (2020). [cited 2023 Jan 29];Available from https://www.cbsnews.com/sanfrancisco/news/livermore-man-arrested-for-1997-sexual-assaults-dna-testing-played-major-role/.

Donaghue, E. (2019). Investigators solve 1970s cold cases with emerging DNA technique. *CBS News*. [cited 2023 Jan 29];Available from https://www.cbsnews.com/news/investigators-solve-1970s-cold-cases-with-emerging-dna-technique-today-2019-02-25/.

Dowd, K. (2020). *Suspect arrested in 32-year-old bay area cold case murder of baby john doe*. SFGATE. [cited 2023 Jan 28];Available from https://www.sfgate.com/crime/article/Suspect-arrested-in-32-year-old-Bay-Area-cold-15455476.php.

Egan, L. (2019). *Mother who stuffed newborn girl 'julie valentine' in box and tossed her woods is also connected to death of newborn son: police*. Crime Online. [cited 2023 Jan 29];Available from https://www.crimeonline.com/2019/05/17/mother-who-stuffed-newborn-girl-julie-valentine-in-box-and-tossed-her-woods-is-also-connected-to-death-of-newborn-son-police/.

Ellis, R. (2018). *DNA on napkin used to crack 32-year-old cold case, police say*. CNN. [cited 2023 Jan 29];Available from https://www.cnn.com/2018/06/22/us/cold-case-killing-1986/index.html.

Emily, S. (1973). *DNA leads to man's arrest 46 years after murder of 11-year-old linda o'keefe: 'we have never forgotten linda'*. [cited 2023 Jan 29];Available from https://abcnews.go.com/US/california-cold-case-solved-50-years-11-year/story?id=61181451.

Erica Jones, S. B. (2018). 'DNA gave us a break:' genealogy tools lead montgomery co-police to cold case rape suspect. *NBC4 Washington*. [cited 2023 Jan 29];Available from https://www.nbcwashington.com/news/local/arrested-made-in-series-of-rapes-in-montgomery-county-police-say/171860/.

Figueroa, T. (2019). *Prosecutor: victim of fatal sword attack in 2006 had been stabbed 30+ times*. Tribune. [cited 2023 Jan 29];Available from https://www.sandiegouniontribune.com/news/courts/sd-me-sword-stabbing-arraignment-20190129-story.html.

Fortin, J. (2018). In serial rape case that stumped police, genealogy database leads to arrest. *The New York Times*. [cited 2023 Jan 29];Available from https://www.nytimes.com/2018/08/23/us/ramsey-street-rapist-dna.html.

Garske, M. (2018). *DNA technology identifies suspect in woman's unsolved valentine's day 2007 killing*. NBC 7 San Diego. [cited 2023 Jan 29];Available from https://www.nbcsandiego.com/news/local/jodine-serrin-carlsbad-cold-case-murder-suspect-david-mabrito/168412/.

Hale-Shelton, D. (2018). *Officials identify suspect in 1990 rape, killing of arkansas woman abducted from video store*. Arkansas Online. [cited 2023 Jan 29];Available from https://www.arkansasonline.com/news/2018/oct/30/officials-identify-suspect-in-1990-rape/.

Halperin, E. (2020). *Delaware county cold case homicide from 1983 solved with help of dna database*. NBC4 WCMH-TV. [cited 2023 Jan 29];Available from https://www.nbc4i.com/news/local-news/delaware-county-cold-case-homicide-from-1983-solved-with-help-of-dna-database/.

Hutton, C. (2019). *Who am i? mystery of a 1978 lynnwood john doe is solved*. HeraldNet.com. [cited 2023 Jan 29];Available from https://www.heraldnet.com/news/who-am-i-mystery-of-a-1978-lynnwood-john-doe-is-solved/.

Hutton, C. (2020). *Solved: for 43 years, she was 'precious' jane doe*. HeraldNet.com. [cited 2023 Jan 29];Available from https://www.heraldnet.com/news/solved-for-43-years-she-was-precious-jane-doe/.

Johnson, G. (2019). 52-year-old Seattle murder case is the oldest ever to be solved using genealogy. *The Seattle Times*. [cited 2023 Jan 29];Available from https://www.seattletimes.com/seattle-news/law-justice/dna-family-tree-help-solve-52-year-old-seattle-slaying/.

Journal, W. M. D. (2018). Rienzi man charged with 1990 Starkville murder. *Northeast Mississippi Daily Journal*. [cited 2023 Jan 29];Available from https://www.djournal.com/news/rienzi-man-charged-with-starkville-murder/article_71089af3-2aac-51f1-90fc-f5b7e04b8069.html.

Kiki Intarasuwan, T. A. P. (2020). *UC berkeley employee arrested in 27-year-old norcal rapist case*. NBC Bay Area. [cited 2023 Jan 29];Available from https://www.nbcbayarea.com/news/california/sacramento-law-enforcement-officials-development-in-norcal-rapist-case/183451/.

Knapp, G. (2019). *Man stolen as a baby found living in michigan 55 years after crime*. KLAS. [cited 2023 Jan 29];Available from https://www.8newsnow.com/news/local-news/man-stolen-as-a-baby-found-living-in-michigan-55-years-after-crime/.

Knox, A. (2020). Serial rapist to spend life in prison for decades-old attacks in northern Utah. *The Deseret News*. [cited 2023 Jan 29];Available from https://www.deseret.com/utah/2020/4/6/21209682/serial-rapist-trucker-utah-wyoming-dna-sentenced-life-in-prison.

Laster, L. (2019). *A look at the technology used in the now closed, 45-year-old double-homicide case out of billings*. ABC FOX Montana. [cited 2023 Jan 29];Available from https://www.montanarightnow.com/regional/a-look-at-the-technology-used-in-the-now-closed/article_2faee207-31c5-5568-95ec-7cb72409659f.html.

Lotan, G. T. (2019). *Orlando police solve 2001 cold case murder using dna, genealogy database*. Orlando Sentinel. [cited 2023 Jan 29];Available from https://www.orlandosentinel.com/news/crime/os-ne-orlando-police-cold-case-murder-arrest-20181105-story.html.

Lotmore, M. (2020). A cigarette leads to an arrest in a 27-year Cold case. *Lynnwood Times*. [cited 2023 Jan 28];Available from https://lynnwoodtimes.com/2020/07/29/a-cigarette-leads-to-an-arrest-in-a-27-year-cold-case/.

Luke, B. (2018). *Indiana man who taunted police pleads guilty to 1988 murder of april tinsley*. NBCNews.com. [cited Jan 29 2023]. Available from https://www.nbcnews.com/news/us-news/indiana-man-who-taunted-police-pleads-guilty-1988-murder-april-n945341. [cited Jan 29 2023]. Available from.

McGee, M. (2019). How genealogists helped track down the Maine man accused of killing Sophie Sergie nearly 26 years ago. *Anchorage Daily News*. [cited 2023 Jan 29];Available from https://www.adn.com/alaska-news/crime-courts/2019/02/17/how-genealogists-helped-track-down-the-maine-man-accused-of-killing-sophie-sergie-nearly-26-years-ago/.

Milkovits, A. (2018). How DNA, a tattoo led to charges in cold murder case. *The Providence Journal*. [cited 2023 Jan 29];Available from https://www.providencejournal.com/story/news/courts/2018/07/27/how-dna-and-tattoo-led-to-charges-in-cold-ri-murder-case/11202328007/.

Munoz, C. R. (2018). *Sheriff: New DNA technology solves 1999 murder of sarasota woman*. Tribune. [cited 2023 Jan 29];Available from https://www.heraldtribune.com/story/news/local/sarasota/2018/09/19/sheriff-new-dna-technology-solves-1999-murder-of-sarasota-woman/10255412007/.

Muskogee man charged with five counts of rape. (2020). Muskogee Phoenix. [cited 2023 Jan 29];Available from https://www.muskogeephoenix.com/news/muskogee-man-charged-with-five-counts-of-rape/article_dc890017-d937-5b27-ad8a-a587f1332ffc.html. [cited 2023 Jan 29];Available from.

Newsmedia. (2020). *Chandler police arrest sex assault suspect in 1991 cold case*. The Daily Independent at Your-Valley.net. [cited 2023 Jan 29];Available from https://www.yourvalley.net/stories/chandler-police-arrest-sex-assault-suspect-in-1991-cold-case,178002.

Newsroom NBC. DNA testing helps id man found near James River in 2016 [cited 29 January 2023]. Available from: https://www.nbc12.com/2019/03/29/dna-testing-helps-id-man-found-near-james-river/.

Petrie, Bonnie (2019). *A genetic "witness" links an oregon cold case to a texas killer*. Texas Public Radio. [cited 2023 Jan 29];Available from https://www.tpr.org/bioscience-medicine/2019-02-05/a-genetic-witness-links-an-oregon-cold-case-to-a-texas-killer. [cited 2023 Jan 29];Available from.

Pomeroy, C. (2019). *COPS: Potomac river rapist, who killed a woman and raped many others in the 90s, arrested*. WJLA. [cited 2023 Jan 29];Available from https://wjla.com/news/local/potomac-river-rapist-arrested.

Queally, James (2019). L.A. prosecutors use genealogy search to make arrest in 1980s killings of two women. *Los Angeles Times*. [cited 2023 Jan 29];Available from https://www.latimes.com/california/story/2019-11-15/cold-case-murders-genealogy-search. [cited 2023 Jan 29];Available from.

Reynolds, W. (2020). *42-year old cold case murder solved; real estate agent's body was found at cantonment home*. NorthEscambia.com. [cited 2023 Jan 28];Available from http://www.northescambia.com/2020/06/42-year-old-cold-case-murder-solved-real-estate-agents-body-was-found-at-cantonment-home.

Ruiz, E. (2020). 40-year-old cold case solved with new genetic genealogy technology. *Denver 7 Colorado News*. [cited 2023 Jan 29];Available from https://www.denver7.com/news/national-politics/the-race-2020/40-year-old-cold-case-solved-with-new-genetic-genealogy-technology.

Sernoffsky, E. (2018). *Suspect in 1973 palo alto killing caught through DNA matchup*. San Francisco Chronicle. [cited 2023 Jan 29];Available from https://www.sfchronicle.com/crime/article/Murder-suspect-in-1973-Palo-Alto-cold-case-caught-13412437.php.

Shapiro, E. (2019a). *ABC news*. [cited 2023 Jan 29];Available from https://abcnews.go.com/US/dna-cigarette-butt-leads-mans-arrest-1994-murder/story?id=62752370.

Shapiro, E. (2019b). DNA links Navy classmate to 1984 cold case murder of 25-year-old Pamela Cahanes: officials. *ABC News*. [cited 2023 Jan 29];Available from https://abcnews.go.com/US/dna-links-navy-classmate-1984-cold-case-murder/story?id=61687544.

Shapiro, E. (2023). DJ admits to cold case murder after being linked through novel DNA technique, apologizes to victim's family in court. *ABC News*. Available from https://abcnews.go.com/US/dj-admits-cold-case-murder-linked-dna-technique/story?id=60234009.

Staff, K. C. R. A. (2022). *SAC sheriff: arrest made in 40-year-old cold case using dna evidence*. [cited 2023 Jan 29]; Available from https://www.kcra.com/article/arrest-made-murder-robin-brooks-sacramento-county/32270681.

Staff CBS17. (2019). *DNA used to charge man in connection with 9 fayetteville cold case rapes*. CBS17.com. [cited 2023 Jan 29];Available from https://www.cbs17.com/news/local-news/cumberland-county-news/dna-used-to-charge-man-in-connection-with-9-fayetteville-cold-case-rapes/.

Staff KIRO7 N. (2019). Edmonds man pleads not guilty in 1972 murder of Jody Loomis. *KIRO 7 News Seattle*. [cited 2023 Jan 29];Available from https://www.kiro7.com/news/north-sound-news/arrest-made-in-1972-murder-of-mill-creek-area-woman/939084943/.

Stelloh, T. (2019). *Indiana police say 47-year-old murder of college student pamela milam solved*. NBCNews.com. [cited 2023 Jan 29];Available from https://www.nbcnews.com/news/us-news/indiana-police-solve-47-year-old-murder-college-student-pamela-n1002596.

The case of Pamela Maurer's murder went cold for decades, then genetic sleuthing by Parabon helped crack it. *CBS News*, (2020). [cited 2023 Jan 29];Available from https://www.cbsnews.com/chicago/news/pamela-maurer-bruce-lindahl-parabon/.

Theisen, T. (2020). 'Valentine Jane Doe homicide' cold case solved 29 years after teen girl found dead in Florida. *Orlando Sentinel*. [cited 2023 Jan 29];Available from https://www.orlandosentinel.com/news/florida/os-ne-florida-keys-rape-murder-cold-case-solved-dna-technology-20200615-cscvcegv7rhylfx5igyuaqvfby-story.html.

Uria, D. (2020). *Callejo, calif., police use DNA to solve 46-year-old cold case killing of naomi sanders*. UPI [cited 2023 Jan 29];Available from: https://www.upi.com/Top_News/US/2020/02/27/California-police-use-genealogy-database-to-solve-46-year-old-cold-case-killing/6311582841581/.

Waldrop, T. (2019). *An arizona man is arrested in the 1979 murder of a woman in nevada*. CNN [cited 2023 Jan 29];Available from: https://www.cnn.com/2019/11/17/us/nevada-arrest-murder-cold-case/index.html.

Wfla. (2019). DNA, genetic genealogy helps investigators crack unsolved rape of 12-year-old. *KSNT 27 News*. [cited 2023 Jan 29];Available from https://www.ksnt.com/news/dna-genetic-genealogy-helps-investigators-crack-unsolved-rape-of-12-year-old/.

Wood, S. (2019). *1974 peninsula cold case linked to hayward man: smc sheriff's*. Los Altos, CA Patch. [cited 2023 Jan 29];Available from https://patch.com/california/losaltos/s/gpub5/1974-peninsula-cold-case-linked-to-convicted-murderer-smc-sher.

Wqow. (2019). *Suspect, victim identified in 20-year-old wisconsin cold case*. WQOW. [cited 2023 Jan 29];Available from https://www.wqow.com/archive/suspect-victim-identified-in-20-year-old-wisconsin-cold-case/article_b08dfecb-6057-5b8b-baee-3a905bd227a8.html.

Wu, G. (2019). *28-year-old fremont cold case solved, thanks to DNA match, police say*. San Francisco Chronicle [cited 2023 Jan 29];Available from: https://www.sfchronicle.com/crime/article/28-year-old-Fremont-cold-case-solved-thanks-to-13559060.php.

Yang, A., & Halaban, B. (2021). Serial killer Terry Rasmussen's victims, known and unknown. *ABC News*. [cited 2023 Jan 29];Available from https://abcnews.go.com/US/terry-rasmussens-victims-unknown/story?id=69585534.

SECTION 3

Challenges in implementing NGS in forensic laboratories

This page intentionally left blank

CHAPTER 24

Troubleshooting and challenges of Next-generation sequencing technology in forensic use

Eugenio D'Orio[1], Aurora Toso[1], Paola Montagna[1] and Maria Denise Gemmellaro[2]
[1]Bio Forensics Research Center, Ischia, Italy; [2]Department of Biology, Kean University, Union, NJ, United States

Forensic genetics: timeline of technology evolution

DNA sequencing was first described by Maxim, Gilbert, and Sanger et al. in 1977 (Maxam & Gilbert, 1977; Sanger et al., 1977). Subsequently, this methodology of analysis evolved into *Next-Generation Sequencing (NGS)* techniques, which, to date, are also known by the term *Massive Parallel Sequencing (MPS).* An MPS technique is defined by the *National Cancer Institute*'s Dictionary of Genetic Terms as: "A high-throughput method used to determine a portion of the nucleotide sequence of an individual's genome." This technique uses DNA sequencing technologies that are capable of processing multiple DNA sequences in parallel.

DNA NGS sequencing techniques have undergone significant developments over the years, which have allowed us to subdivide these analysis methodologies into *first*, *second*, *and third generation sequencing*. A variety of important developments in first, second, and third generation sequencing are depicted in Fig. 24.1.

As can be seen from the timeline, it is clear that the first-generation *Sanger* sequencer, developed in 1977, dominated the market for a long time. In the years 2005—2007 several second generation systems were launched in the market, such as the *Solexa Genome Analyzer,* from Illumina, and the *SOLiD R* system from Applied Biosystem. The so-called third generation sequencers started with the launch of the Helicos Genetic Analysis Platform in 2007(from Helicos BioSciences) that was the first DNA-sequencing instrument to operate by imaging individual DNA molecules. In 2014, Oxford Nanopore Technologies (ONT) MinION, which was introduced for early access, is a pocket-sized device that applies nanopore sequencing technology to nucleic acid analysis.

The scaled-up GridION was commercially launched in 2017 and PromethION in 2018, with the largest device, the PromethION 48, first shipped in 2019 (Nanopore sequencing-Oxford Science Park, Oxford, UK). All of these technologies that have developed over the years since the Sanger method have increased efficiency and accuracy by more than three orders of magnitude (Bruijns et al., 2018).

In principle, the concepts behind Sanger versus next-generation sequencing (NGS) technologies are similar. In both NGS and Sanger sequencing, DNA polymerase adds

Next Generation Sequencing (NGS) Technology in DNA Analysis
ISBN 978-0-323-99144-5, https://doi.org/10.1016/B978-0-323-99144-5.00006-8
© 2024 Elsevier Inc.
All rights reserved.

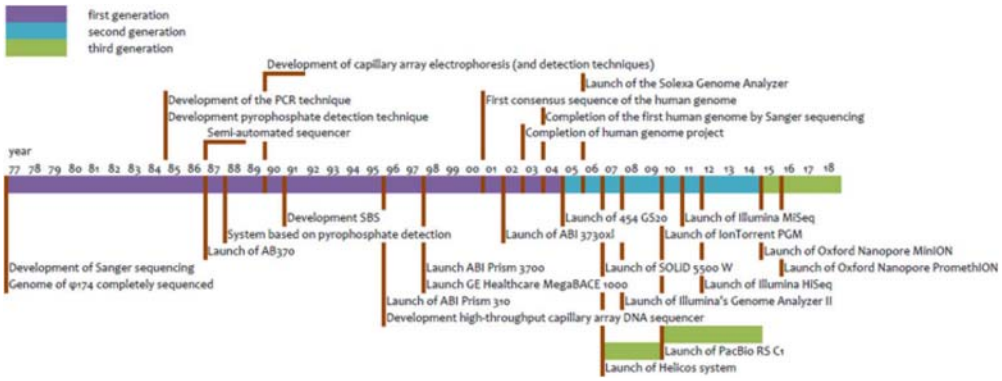

first generation
second generation
third generation
Development of capillary array electrophoresis (and detection techniques)
Launch of the Solexa Genome Analyzer
Development of the PCR technique
Development pyrophosphate detection technique
First consensus sequence of the human genome
Completion of the first human genome by Sanger sequencing
Semi-automated sequencer
Completion of human genome project
year
77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 00 01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16 17 18
Development SBS
Launch of 454 GS20
Launch of Illumina MiSeq
System based on pyrophosphate detection
Launch of ABI 3730xl
Launch of IonTorrent PGM
Launch of AB370
Launch of Oxford Nanopore MinION
Development of Sanger sequencing
Launch ABI Prism 3700
Launch of SOLID 5500 W
Launch of Oxford Nanopore PromethION
Genome of φ174 completely sequenced
Launch GE Healthcare MegaBACE 1000
Launch of Illumina HiSeq
Launch of ABI Prism 310
Launch of Illumina's Genome Analyzer II
Development high-throughput capillary array DNA sequencer
Launch of PacBio RS C1
Launch of Helicos system

Figure 24.1 Timeline of the most important developments regarding NGS. *(From Bruijns, B., Tiggelaar, R., Gardeniers, H. (2018). Massively parallel sequencing techniques for forensics: A review.* Electrophoresis, *39(21), 2642–2654. doi: 10.1002/elps.201800082. Epub 2018 Aug 22. PMID: 30101986; PMCID: PMC6282972.)*

fluorescent nucleotides one by one onto a growing DNA template strand. Each incorporated nucleotide is identified by its fluorescent tag. The critical difference between Sanger sequencing and NGS is sequencing volume. While the Sanger method only sequences a single DNA fragment at a time, NGS is massively parallel, sequencing millions of fragments simultaneously per run. This process translates into sequencing hundreds to thousands of genes at one time.

With each step, more sophisticated bioinformatics tools, programs, and software for DNA sequencing have provided more automation and higher throughput. In recent years, several parallel sequencing (NGS) methods have become available, allowing large-scale production of genomic sequence, and the number of human genomes sequenced with such instrumentation has increased rapidly.

First generation sequencing

Sanger sequencing was developed in 1977 by Frederick Sanger, who was awarded the Nobel Prize in Chemistry in 1980 (Sanger et al., 1977). This method, which is now known as the first-generation sequencing, is rather similar to PCR, because an ssDNA template, a DNA primer, DNA polymerase, and deoxynucleotide triphosphates (dNTPs) are also required to perform the reaction. Furthermore, di-deoxynucleotide triphosphates (ddNTPs) are needed, which can be incorporated into the newly synthesized DNA strand, just as normal dNTPs, but with termination of the elongation process as a result. Therefore, this method is also known as the chain termination method. The reaction is carried out in fourfold, whereby in each tube (besides the normal dNTPs) only one of the four labeled ddNTPs is added in a relatively low concentration. Nowadays the sequence can be determined by fluorescently labeled ddNTPs and capillary (gel) electrophoresis (ibidem).

Second generation sequencing

The development of PCR in 1985 has led to major improvements in instrumentation, because alternatives for the principle of Sanger sequencing became available. The new systems such as *454*, *SOLiD R*, and *Ion TorrentTM* make use of a cell-free system, whereas for Sanger sequencing bacterial cloning of DNA fragments was required (Kumar, 2012; Shendure & Ji, 2008; Van Dijk et al., 2014).

Compared to Sanger sequencing, the second generation NGS techniques allow us to obtain a huge amount of data (sequences) in less time and with lower costs. This allows us to analyse multiple targets at the same time (up to the entire genome of an organism) of the same sample, analyse multiple samples at the same time for the same targets or even for different targets between samples, produce millions of sequencing reactions in parallel compared to tens or hundreds of Sanger and not to requery the bacterial cloning of DNA fragments as it is based on cell-free systems for the preparation of libraries.

There are different types of platforms (e.g., Roche, Thermo Fisher Scientific, Illumina) each with different tools (e.g., MiSeq and HiSeq for Illumina or SOLID and Ion Torrent for Thermo Fisher Scientific). Each has its own distinctive features, but they all share the commonality of using DNA or cDNA (RNA) libraries as input for sequencing. Two primary MPS platforms are used in forensic DNA analysis: (1) MiSeq FGx Forensic Genomics Systems (Illumina, San Diego, CA, USA) and (2) Ion Tor-rent PGM or Ion S5 (ThermoFisher Scientific, Waltham, MA, USA) (Haddrill, 2021).

Third generation sequencing

Although second generation sequencing techniques are based on amplification, with the next-next-next (or third generation) methods single molecules are read in real time. Therefore, these techniques are much faster and longer reads can be generated then with the previous generations of sequencing techniques. Single molecule sequencing (SMS) technologies can be grouped in three categories. The first category is SBS methods, whereby single molecules of DNA polymerase are observed at the moment they are synthesizing a single DNA molecule (e.g., PacBio R). Nanopore sequencing techniques are the second category (e.g., Oxford Nanopores) and the third category consists of methods that use direct imaging of individual DNA molecules by means of advanced microscopy (e.g., VisiGen) (Bruijns et al., 2018).

Current limitations and problem solving in forensic genetics

To date, the technology used for short tandem repeat (STR) marker typing is based on multiplex PCR followed by capillary electrophoresis and is well established in all forensic genetics' laboratories. In contrast, MPS platforms require a different analytical workflow and new forensic bioinformatics resources that are currently not widely available due to the enormous amount of data generated. In fact, it must be considered that millions of

reads per sample can be obtained from a single sequencing experiment conducted with MPS methodology. Traditional STR analysis by gel electrophoresis or *capillary electrophoresis* (CE) estimates the number of STR repeats by the size of the PCR amplicon. In contrast, analysis with NGS allows determination of the entire sequence of the PCR product, including the STR repeat region and surrounding flanking areas. Numerous studies have shown that sequence-based STR analysis, as opposed to amplicon size-based analysis, results in a pronounced increase in allelic variability for many forensically relevant STRs (Gettings et al., 2017). Even loci that appear to show no additional sequence variation in individual populations can be informative in others, as demonstrated by a study of four major population groups in the USA where only TPOX of the commonly used autosomal loci failed to show added sequence variation in any of the investigated populations (Phillips et al., 2018). Capturing the additional sequence variation that may be present in the STR repeat and flanking regions can have many advantages beyond simply improving discrimination power when using these markers for direct matching or relationship calculations. For DNA mixtures, the improved discrimination of the markers will in itself be very advantageous for identifying extra sequence-specific alleles that would have been previously masked due to identical CE amplicon lengths, while the enhanced discrimination will also make it less likely that an individual will have an adventitious match between their profile and the alleles present within a mixed trace.

One notable advantage of analysis by MPS rather than CE is that separation by amplicon size is no longer a requirement for multiplex STR assay design; this means that all STRs can be amplified using the smallest feasible amplicon length, improving amplification in cases of degraded DNA, and that the number of markers amplified simultaneously is no longer constrained by the fluorescent dye detection capabilities of electrophoretic systems hence heralding the possibility of coamplifying increasing numbers of STRs at the same time. One use of this would be to allow coamplification of large panels of Y and autosomal STRs, providing increased power for male/female mixed stain analysis, e.g., following sexual assault (Ballard et al., 2020).

However, the amplification of forensic DNA samples by MPS is not limited to *STR loci* though, and *SNP markers* can also be analysed either in combination with STRs or on their own. The advantages of analysing SNPs for forensic samples with MPS rather than alternative technologies such *as SNaPshot* or *SNP array systems*, rest predominantly in the fact that it is possible to target large (compared with historical SNP systems used in forensic genetics) numbers of markers simultaneously from low quantities of DNA. It is also possible to target microhaplotypes with NGS, that is, sets of SNPs located in very close proximity to each other on a chromosome, and these microhaplotype markers have shown promise in forensics for both identification and ancestry purposes (Wendt et al., 2017). Moreover, the fact that the entire PCR amplicons are analysed when typing SNP markers with NGS means that any observed variation in the PCR flanking regions will also in effect turn

the targeted SNP into a microhaplotype. One example of this is reported by Wendt et al. (2017). when studying a Native American population; they found that 22 of the 94 identity SNPs studied contained flanking region variants in the tested population of which 14 were informative (i.e., not in total linkage with one specific allele) the fact that some identity SNPs are actually microhaplotypes is to be expected and the additional information that this extra variation adds will be valuable for all applications.

Another important factor to consider regarding NGS methodology is that currently there are a few NGS bench-top sequencers that dominate the forensic genetics landscape: *Illumina's MiSeq FGx*, *ThermoFisher's Ion Torrent PGM*, and *Ion S5*. While *Illumina* implements cycle-based sequencing technology coupled with a reversible termination strategy of fluorescently labeled modified dNTPs, conceptually similar to Sanger sequencing, the *Ion Torrent* is a semiconductor sequencer that measures pH changes as consequence of the release of hydrogen ions during synthesis of DNA. While base substitutions are the most common sequencing errors generated during sequencing on Illumina machines, insertions and deletions are the most frequent errors introduced by the Ion Torrent PGM. In the case of the latter, homopolymer stretches longer than 6 bp are the most difficult to call, since the correlation between the incorporated nucleotides and the change in detected voltage is not exactly to scale. Although the error rates generated by the Ion Torrent are higher ($\geq$1%) and DNA library preparation protocols can be more time-consuming and cumbersome in comparison with the MiSeq workflow, the lack of optical scanning and cycle-based sequencing significantly reduces the time of DNA sequencing. One notable advantage of the Ion Torrent platforms is the availability of automated library preparation and chip loading stations, which simplifies the workflow considerably. Recently, similar automation and liquid handling solutions have been proposed for many MiSeq workflows (Laurent et al., 2017).

Among the existing commercial solutions the *ForenSeq DNA Signature Prep Kit* (Illumina, CA) is the first commercially available STR kit for the *MiSeq FGx* that allows for amplification of up to 153 (DNA primer mix A) or 231 (DNA primer mix B) loci simultaneously. The multiplex assay includes 27 forensic autosomal STRs, 24 Y-STRs, 7-X STRs, the Amelogenin sex marker, and either 94 or 172 SNPs depending on the multiplex formulation. The 172 SNPs can be divided into informative identity, geographical ancestry, and phenotypic SNPs (Ballard et al., 2020).

NGS: from diagnostics to forensic applications

Next-generation sequencing (NGS) has enabled scientists and clinicians to better understand the genetic mechanisms associated with disease conditions that lead to altered health status and the development of serious diseases such as cancer and infectious illnesses. To date, NGS has become essential in application areas such as: medical diagnostics, forensic genetics, biotechnology, and virology.

These advances have resulted in improved diagnostics in the clinic for early intervention and monitoring of treatment response, ensuring patients receive the best possible therapies.

Use of NGS in clinical diagnosis is now widely accepted, with varying roles from gene panel or targeted sequencing, through *whole exome sequencing* (also termed clinical exome sequencing), to *whole genome sequencing*. NGS was first applied to clinical diagnosis when clinical exome sequencing was launched in 2012. The neurology community, although recognizing its technical limitations, including nonuniform coverage and the difficulty in sequencing or analyzing certain genomic regions and lack of ease in detecting certain types of variants, still believed the new technology had benefits to offer. The first scientific reports showed that *clinical exome sequencing* was able not only to successfully identify both rare and common single nucleotide variants and small insertions and deletions across coding exons and flanking splice sites of genes, but also to achieve a higher diagnostic rate than most of the traditional molecular tests, such as single gene sequencing, small gene panels, or chromosomal microarrays for rare Mendelian disorders.

Furthermore, diagnostic rates between laboratories were surprisingly consistent despite differences in the capture kits used, bioinformatics pipelines, and patient cohorts. Initially, variant confirmation was done with Sanger sequencing, which was, and still is in some estimations, considered the gold standard. However, laboratories soon began to develop internal quality measurements to determine the necessity of confirmation for each variant.

We now know that there are genomic regions, and conditions in which variant abundance is relevant to disease, for which NGS is a better method than Sanger sequencing, raising the question of whether Sanger sequencing should be replaced as the gold standard (Haddrill, 2021).

Additionally, the *Next-generation sequencing (NGS*) methods effectively allow all types of nucleic acids to be sequenced, using a whole genome or targeted approach, with *DNA, mRNA*, and *small RNA* sequencing as standard analyses. Furthermore, larger-scale sequencing of specific RNA subtypes, such as long noncoding RNAs and snoRNA, as well as methylated DNA, have become possible with the introduction of NGS. The prospect of simultaneously analyzing a large number of markers such as STRs and SNPs in parallel with targeted mRNA and small RNA analysis makes MPS a very powerful, relatively easily applicable, tool in forensic laboratories (Fig. 24.2) (Ballard et al., 2020).

In particular, the relevance to forensic analysis is that in addition to generating standard *STR* profiles, DNA repeats can also be sequenced to look for *single nucleotide polymorphisms* (*SNPs*).

Moreover, additional SNPs can be sequenced to acquire ancestry, paternity, or phenotype information.

Current MPS systems are also very useful in cases where DNA has been acquired from a crime scene so it may be available in limited quantities or is present in a highly

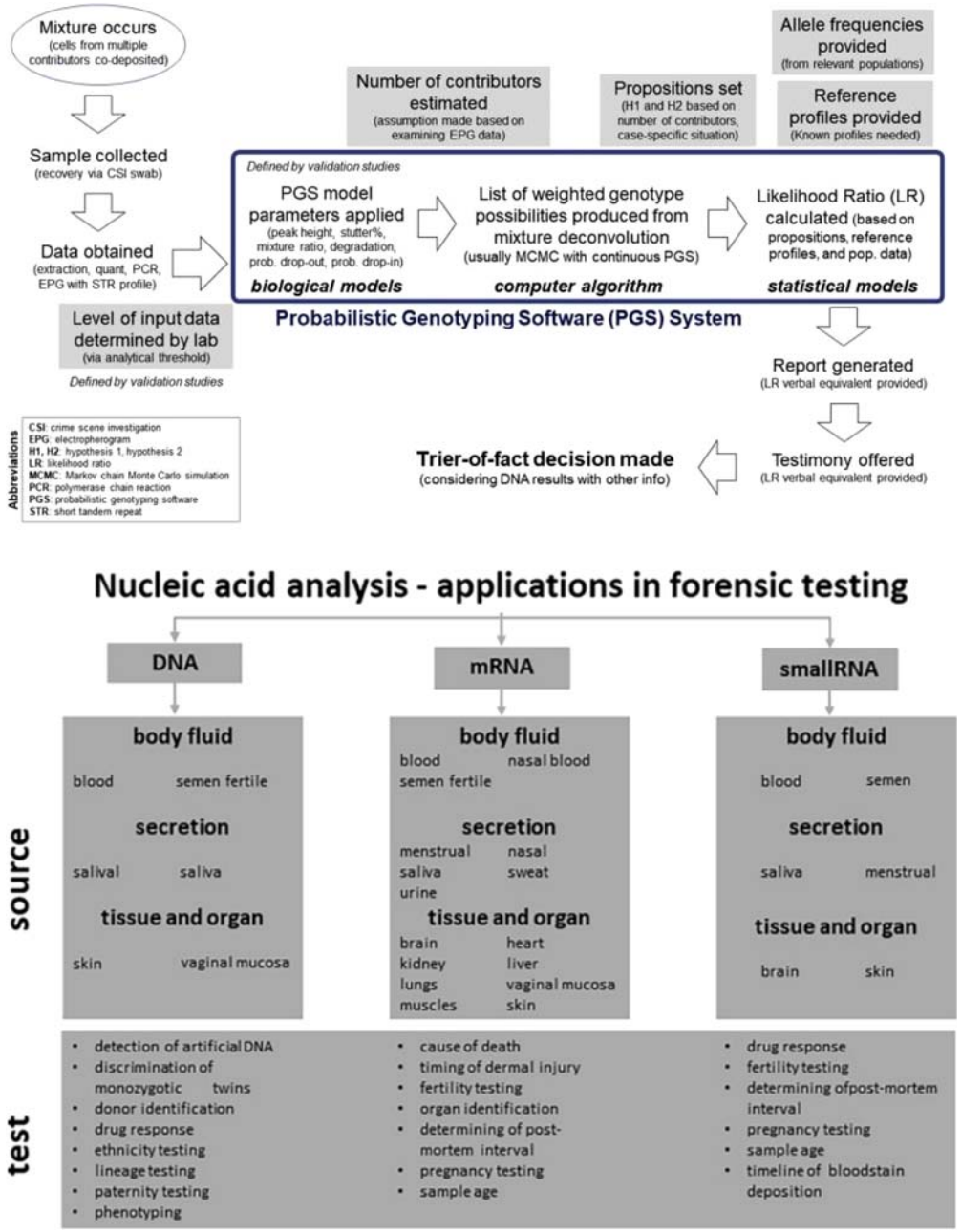


Figure 24.2 Most prevalent applications of nucleic acid analysis in forensic testing. *(From Ballard, D., Winkler-Galicki, J., Wesoły, J. (2020). Massive parallel sequencing in forensics: advantages, is-sues, technicalities, and prospects.* International Journal of Legal Medicine, 134*(4), 1291—1303. doi: 10.1007/s00414-020-02294-0. Epub 2020 May 25. PMID: 32451905; PMCID: PMC7295846.)*

degraded form. If, for example, not enough autosomal DNA is present and a lineage analysis needs to be performed, *mitochondrial DNA* can be sequenced, via NGS, for maternal lineage analysis. The application of MPS technology also extends to other markers such as DNA methylation patterns and mRNA gene expression for identification of body fluids.

The use of this wide variety of markers is useful for *forensic DNA phenotyping* (FDP) and DNA-based prediction of appearance (eye, hair, and skin color), biogeographic ancestry, and age. For this reason, the "*Visible Attributes through Genomics - VISAGE*" (Gross et al., 2021) consortium is developing MPS tools for FDP marker analysis to enable prediction of appearance, ancestry, and age of unidentified contributors from biological stains. These tools are designed to allow forensic laboratories to implement these innovative approaches into their repertory of routine applications for studying case samples.

These developments clearly demonstrate that the use of NGS will become an indispensable tool for forensic science, and the various components present in a mixture and assigning an appropriate weight to the evidence can be challenging (Butler & Willis, 2020). The complexity of mixed profiles has called for increasingly complex methods of mixture interpretation, and there has been a move away from relatively simple methods that ranged from determining whether an individual could be excluded as a potential contributor to a mixture, to the use of likelihood ratio methods that estimated the most likely genotype combinations of contributors to a mixture, the more complex of which used some of the information contained within profile peak heights (Haddrill, 2021).

This has led over the years to an increase in the use of *probabilistic genotyping software (PGS)* to assist DNA mixture interpretation. Generally, PGS systems use either: (1) "discrete" (sometimes called "semi-continuous") models that use the presence or absence of peaks along with probabilities of allele drop-out or drop-in or (2) "continuous" (sometimes called "fully-continuous") models that take peak heights into account as well as the presence or absence of peaks along with probabilities of allele drop-out or drop-in. Fig. 24.2 describes the general steps in mixture interpretation along with user inputs required for PGS systems. The ability of these programs to analyse mixtures previously considered too complicated for interpretation has seen rapid uptake by forensic laboratories, and publication of studies reporting the developmental and internal validation of different probabilistic genotyping software packages, as well as guidelines for their use by a number of regulating bodies (Forensic Science Regulator, 2020; Haddrill, 2021). The software packages that implement probabilistic genotyping methods are highly complex, and developers have urged forensic laboratories to ensure their analysts have a good understanding of the concepts underlying the methods and that they remain involved in the interpretation of profiles and critical evaluation of the mixture analysis (Lee et al., 2019). Concerns have been raised over variation in

the output of probabilistic genotyping methods, some due to subjective decisions made by the user, some due to variability inherent in the methods (Haddrill, 2021). Some countries have seen extensive debates over the admissibility of probabilistic genotyping methods in court and whether methods have gained general acceptance in the community, but the widespread implementation of these methods into forensic laboratories around the world suggests that the response has been positive. These methods also provide significant promise for the interpretation of MPS data that can uncover greater complexity in mixed profiles by identifying sequence differences between alleles that would be indistinguishable by length.

NGS in forensics and the issue of privacy: a global overview

The development over the years of increasingly cutting-edge NGS techniques has made forensic DNA testing a significant resource for investigation and prosecution activities in criminal justice systems around the world. Forensic DNA testing can be conducted in several ways: first, by comparing the DNA profiles from criminal suspects to DNA evidence, so as to assess the likelihood of their involvement in a crime. The second kind of use is related to searching for a link between the biological material collected from a crime scene and a DNA profile stored in a criminal DNA database. The third form of forensic DNA testing is related to procedures to search for criminal suspects through their connection with biological relatives. Finally, the inference of human externally visible physical features from a biological sample collected at the crime scene (Toom, 2018). One prominent aspect of forensic DNA testing is the establishment and expansion of centralized national criminal DNA databases. These databases involve the collection, storage, and use of DNA profiles from nominated suspects, convicted offenders, victims, volunteers, and other persons of interest to criminal investigation work. The primary function of a criminal DNA database is to produce matches between individual profiles and crime scene stains, which requires a constant input of both. Around 69 countries currently operate national forensic DNA databases, and others are being expanded or established in at least 34 additional countries (Interpol, 2016). Recent innovations and developments in forensic DNA testing in the criminal field are related to techniques of Forensic DNA Phenotyping (FDP), the use of ancestry-informative markers, and familial searching. *FDP* can be described as a set of techniques that aims to infer human externally visible physical features—eye, hair, and skin color—and continental-based biogeographical ancestry of criminal suspects on the basis of analysis of biological materials collected at crime scenes. FDP techniques have been applied in various jurisdictions in a limited number of high-profile cases to provide intelligence for criminal investigation. *Familial searching* makes use of procedures to detect genetic relatedness in criminal DNA databases to search for criminal suspects through their connection with biological relatives (Machado & Silva, 2019).

Massively parallel sequencing (MPS), makes possible, in one DNA reaction, the collection of autosomal STRs, Y-STRs, mitochondrial markers, and mRNA as well as a variety of *EVC* (externally visible characteristics) and *AIM* (ancestry informative markers) relevant SNPs and "Indels" (or insertions/deletions). More SNPs may also be included in the arrays that comprise commercially available MPS kits. It is possible to run reactions for STR identification separately, but it is also possible to use the whole array if required. *Genealogical connections* are already routinely interrogated by the use of Y-STRs and mitochondrial DNA, but the inclusion of these in MPS kits may make their uses more common, although the *privacy implications* of this increase are insignificant when confined to the comparison of DNA collected in the course of criminal investigations. Additional considerations may arise if crime-scene DNA is searched against commercially managed "recreational" or research genealogical databases (Williams & Wienroth, 2017) (such as GedMatch or FamilyTreeDNA).

What has just been said can be explained by the fact that, since DNA is sensitive data, in order to guarantee and protect the right to privacy of every citizen, the police cannot use the databases present in genealogy societies on their own initiative unless the citizen, who donates his or her DNA to the database, consents to its use in this specific context.

For this very reason in Maryland, a compromise bill was reached that strikes a balance between this very important technology for identifying people who commit serious crimes and the related privacy concerns. This condition puts a dampener on the use of forensic genealogy in investigations because: first, the bill states that by 2024, genealogists working on such cases must be professionally certified; second, investigators can only use genealogy companies that have explicitly informed the public and their clients that law enforcement uses their databases and have asked for their clients' consent to participate. Currently, GEDmatch and FamilyTreeDNA clients can choose to participate in these searches. Finally, investigators may not use any of the genetic information collected, whether by the suspect or a third party, to learn about a person's psychological traits or predispositions to disease. Upon completion of the investigation, all genetic and genealogical records that were created for it must be deleted from the databases. It has therefore been demonstrated, on the one hand, its effectiveness in terms of providing valuable information for investigations and for the solution of cases, but on the other hand also the strong ethical and legal implications related to privacy.

Different views on the capabilities, benefits, and risks of forensic DNA testing circulate within modern societies. Supporters of the expansion of forensic DNA testing in the criminal justice system invoke its capacity to serve as a valuable law enforcement tool, namely by improving efficiency in fighting crime, helping in the prevention of miscarriages of justice and deterrence of criminal activity, which is, in turn, expected to reduce crime and increase public safety and security. Critics concerned with potential threats to civil liberties argue that forensic DNA testing, in particular the storage of profiles in computerized databases operating as forensic DNA databases for criminal identification,

may threaten the protection of a range of human rights, in particular liberty, autonomy, privacy, informed consent, moral and physical integrity, and the presumption of innocence (Machado e Silva, 2019).

Applications of NGS technology in forensic entomology

Forensic entomology is a branch of the forensic sciences focused on the application of the study of arthropods for legal investigations (Byrd & Tomberlin, 2019); medicolegal entomology, more specifically, focuses on the study of those entomological groups which are attracted to decomposing matter (such as a dead body) and which can provide information useful to the investigators, relative, for example, to postmortem movement of the remains or minimum time since death (Amendt et al., 2011). For the purpose of this chapter, we will focus on the insects of medicolegal interest.

During a criminal investigation that involves entomological evidence, the specimens have to be collected and preserved promptly and correctly, to ensure a correct assessment of their developmental stage, as well as to preserve their morphology. This will make it possible to obtain a correct identification to the lowest taxonomical level of the specimen (ideally species). Being insects poikilothermic creatures, their developmental rate depends on the external temperature (Charabidze & Hedouin, 2019), and it is species-specific. This means that different species develop at different rates at a given temperature. Therefore, to appropriately use insects as indicators of time of colonization (TOC) or minimum postmortem interval (mPMI) (Tomberlin et al., 2011) the determination of the species is essential. Morphological identification is a common method to identify the specimen based on their observable morphological features; this is accomplished through the use of dichotomous keys. Dichotomous keys consist of a specific sequence of couplets referring to a specific morphological trait, which offers two contrasting options; as one proceeds through the couplets, choosing options that apply to the specimen of interest, the specimens are identified through a process of elimination (Charabidze & Martín-Vega, 2021). For some specimens however, morphological identification is not a suitable option because of their conditions (damaged specimens) or because dichotomous keys are not available (Gemmellaro et al., 2019). In those cases, identification has to be accomplished in different ways.

In the field of forensic entomology, molecular biology has been majorly employed to identify entomological specimens (Stevens et al., 2019). The group to which molecular identification has been amply applied is the blowfly family (Diptera: Calliphoridae). Blowflies are among the first colonizers of decomposing remains; the adults can locate a corpse within minutes, and once they arrive on it, they will lay their eggs on it. When these hatch, larvae will eclose and will start feeding on the body voraciously (Greenberg & Kunich, 2002). DNA barcoding is a method consisting in sequencing a short fragment of DNA and in using this fragment to identify the species it belongs to

by comparing it against known sequences stored in DNA libraries. While the term barcoding refers to the process of sequencing one single specimen at the time, metabarcoding refers to massive parallel sequencing of samples containing more than one individual's DNA (Cristescu, 2014). This method has been used extensively for the identification of specimens of forensic interest.

It is possible to use different sequences to identify the species of a specimen; ITS2 (Internal Transcribed Spacer) sequence is one example. This is spacer sequence located between two subunits of ribosomal RNA (rRNA) and it has been vastly used in taxonomy and phylogeny for several taxa, including insects of forensic interest (Gemmellaro et al., 2019). However, most of the molecular identification methods used in forensic entomology for the identification of the specimens focus on the COI + II (Cytochrome *c* Oxidase subunits I and II) mitochondrial genes; these are relatively stable haploid genes whose sequence (entirely or partially) has been used for the species identification of calliphorid flies (Stevens et al., 2019). The BOLD database (Barcode of Life Data System) is a free public resource where barcode COI sequences of known species are stored (Ratnasingham & Hebert, 2007). Currently (2022) the BOLD database contains 12,126 records for Calliphoridae, and a total of 328 different species (https://www.boldsystems.org). Potentially, a specimen recovered on a crime scene that cannot be identified morphologically could undergo DNA isolation, PCR, purification, and sequencing of the COI region. The sequence could then be uploaded on the BOLD database where it would be compared against the known sequences of the library, and an identification to the lowest taxonomical level could be possible.

The revolution represented by the NGS has had a significant impact even in the field of forensic entomology. NGS has allowed for the sequencing of genomic DNA without the need to know the genome of a specific species. Through the use of NGS, it is possible to use DNA fragments to identify fragment of specimens, specimens that have been damaged due to the weather conditions a body may have been on, or even specimens that are too small to be seen and collected (Bittleston et al., 2016). It is common to recover remains colonized by masses of arthropods that cannot be easily separated into individual specimens; depending on the conditions of the remains, there may be just fragments of the arthropods (Chimeno et al., 2019). Without a metabarcoding analysis, the identification of these specimens would likely not be possible.

Another application of NGS in forensic entomology was presented by Farncombe et al. (2014), and it involves the use of NGS to screen a partial genomic sequence of the Black Blow Fly (*Phormia regina* Meigein (Diptera:Calliphoridae)) searching for microsatellite loci; this had the purpose to assess the genetic structure of the population and use this information to infer corpse movements based on the genotypes of the maggots collected on them.

A particular case where NGS analysis was used in forensic entomology but not for the determination of the arthropod species was presented by Pilli et al. (2016). The authors

used NGS techniques on human blood extracted from the gastrointestinal tract of lice and were able to build a clean human profile that was a perfect match with their reference sample. One of the newest applications of NGS to forensic investigations involving entomofauna is its application to the estimation of the postmortem submersion interval (PMSI) or time elapsed between when a body is submerged and its recovery. Hyun et al. (2019) analysed the microeukaryotic biodiversity of a pig carcass and a car bonnet after their submersion in a reservoir; they compared the relative abundance pattern of certain taxa during each of the decomposition stages of the pig to assess potential differences that could potentially be used as references to infer the PMSI.

References

Amendt, J., Richards, C. S., Campobasso, C. P., Zehner, R., & Hall, M. J. (2011). Forensic entomology: Applications and limitations. *Forensic Science, Medicine and Pathology, 7*(4), 379–392.

Ballard, D., Winkler-Galicki, J., & Wesoły, J. (July 2020). Massive parallel sequencing in forensics: Advantages, issues, technicalities, and prospects. *International Journal of Legal Medicine, 134*(4), 1291–1303. https://doi.org/10.1007/s00414-020-02294-0. Epub 2020 May 25. PMID: 32451905; PMCID: PMC7295846.

Bittleston, L. S., Baker, C. C., Strominger, L. B., Pringle, A., & Pierce, N. E. (2016). Metabarcoding as a tool for investigating arthropod diversity in N epenthes pitcher plants. *Austral Ecology, 41*(2), 120–132.

Bruijns, B., Tiggelaar, R., & Gardeniers, H. (November 2018). Massively parallel sequencing techniques for forensics: A review. *Electrophoresis, 39*(21), 2642–2654. https://doi.org/10.1002/elps.201800082. Epub 2018 Aug 22. PMID: 30101986; PMCID: PMC6282972.

Butler, J. M., & Willis, S. (February 20, 2020). Interpol review of forensic biology and forensic DNA typing 2016-2019. *Forensic Science International Synergy, 2*, 352–367. https://doi.org/10.1016/j.fsisyn.2019.12.002. PMID: 33385135; PMCID: PMC7770417.

Byrd, J. H., & Tomberlin, J. K. (Eds.). (2019). *Forensic entomology: The utility of arthropods in legal investigations*. CRC press.

Charabidze, D., & Hedouin, V. (2019). Temperature: The weak point of forensic entomology. *International Journal of Legal Medicine, 133*(2), 633–639.

Charabidze, D., & Martín-Vega, D. (2021). Looking back to move forward: How review articles could boost forensic entomology. *Insects, 12*(7), 648.

Chimeno, C., Morinière, J., Podhorna, J., Hardulak, L., Hausmann, A., Reckel, F., Grunwald, J. E., Penning, R., & Haszprunar, G. (2019). DNA barcoding in forensic entomology—Establishing a DNA reference library of potentially forensic relevant arthropod species. *Journal of Forensic Sciences, 64*(2), 593–601.

Cristescu, M. E. (2014). From barcoding single individuals to metabarcoding biological communities: Towards an integrative approach to the study of global biodiversity. *Trends in Ecology and Evolution, 29*(10), 566–571.

Farncombe, K. M., Beresford, D., & Kyle, C. J. (2014). Characterization of microsatellite loci in *Phormia regina* towards expanding molecular applications in forensic entomology. *Forensic Science International, 240*, 122–125.

Forensic Science Regulator. (2020). *Software validation for DNA mixture interpretation*. FSR-G-223, Issue 2 https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/917784/G223_Mix_software_valid_Issue2_accessV3.pdf.

Gemmellaro, M. D., Hamilton, G. C., & Ware, J. L. (2019). Review of molecular identification techniques for forensically important Diptera. *Journal of Medical Entomology, 56*(4), 887–902.

Gettings, K. B., Borsuk, L. A., Ballard, D., Bodner, M., Budowle, B., Devesse, L., King, J., Parson, W., Phillips, C., & Vallone, P. M. (2017). STRSeq: A catalog of sequence diversity at human identification short tandem repeat loci. *Forensic Science International Genetics, 31*, 111–117.

Greenberg, B., & Kunich, J. C. (2002). *Entomology and the law: Flies as forensic indicators*. Cambridge University Press.

Gross, T., Fleckhaus, J., & Schneider, P. (2021). Progress in the implementation of massively parallel sequencing for forensic genetics: Results of a European-wide survey among professional users. *International Journal of Legal Medicine*. https://doi.org/10.1007/s00414-021-02569-0

https://www.ncbi.nlm.nih.gov/books/NBK367608/#:~:text=Next%2Dgeneration%20sequencing%20(NGS),multiple%20DNA%20sequences%20in%20parallel.

Haddrill, P. R. (September 24, 2021). Developments in forensic DNA analysis. *Emering Top Life Science, 5*(3), 381–393. https://doi.org/10.1042/ETLS20200304. PMID: 33792660; PMCID: PMC8457771.

Hyun, C. H., Kim, H., Ryu, S., & Kim, W. (2019). Preliminary study on microeukaryotic community analysis using NGS technology to determine postmortem submersion interval (PMSI) in the drowned pig. *Journal of Microbiology, 57*(11), 1003–1011.

Interpol. (2016). *Risultati dell'indagine sul profilo del DNA globale*. Accesso il 22 febbraio 2019 https://www.interpol.int/INTERPOL-expertise/Forensics/DNA/Media-library/INTERPOL-Global-DNA-Survey-Results-2016.

Kumar, B. R. (2012). In A. Munshi (Ed.), *DNA sequencing – methods and applications* (pp. 3–14). Rijeka: InTech.

Laurent, F. X., Ausset, L., Clot, M., Jullien, S., Chantrel, Y., Hollard, C., & Pene, L. (2017). Automation of library preparation using Illumina ForenSeq kit for routine sequencing of casework samples. *Forensic Science International: Genetics Supplement Series, 6*, e415–e417.

Lee, H. M., Rexach, J. A., Fogel, J., & Brent, L. (2019). *Next generation sequencing in clinical diagnosis*. https://doi.org/10.1016/S1474-4422(19)30110-3. Crossref DOI link:.

Machado, H., & Silva, S. (2019). What influences public views on forensic DNA testing in the criminal field? A scoping review of quantitative evidence. *Human Genomics, 13*, 23. https://doi.org/10.1186/s40246-019-0207-5

Maxam, A. M., & Gilbert, W. (1977). A new method for sequencing DNA. *Proceedings of the National Academy of Sciences of the United States of America, 74*(2), 560–564. https://doi.org/10.1073/pnas.74.2.560

Phillips, C., Gettings, K. B., King, J. L., Ballard, D., Bodner, M., Borsuk, L., & Parson, W. (2018). The devil's in the detail: release of an expanded, enhanced and dynamically revised forensic STR sequence guide. *Forensic Science International Genetics, 34*, 162–169.

Pilli, E., Agostino, A., Vergani, D., Salata, E., Ciuna, I., Berti, A., Caramelli, D., & Lambiase, S. (2016). Human identification by lice: A next generation sequencing challenge. *Forensic Science International, 266*, e71–e78.

Ratnasingham, S., & Hebert, P. D. (2007). Bold: The barcode of Life data system. *Molecular Ecology Notes, 7*(3), 355–364. http://www.barcodinglife.org.

Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Biotechnology 1992, 24*, 104–108.

Shendure, J., & Ji, H. (2008). *Nature Biotechnology, 26*, 1135–1145.

Stevens, J. R., Picard, C. J., & Wells, J. D. (2019). *Molecular genetic methods for forensic entomology*. Forensic entomology (pp. 253–268). The Utility of Arthropods in Leg. Investigations.

Tomberlin, J. K., Mohr, R., Benbow, M. E., Tarone, A. M., & Vanlaerhoven, S. (2011). A roadmap for bridging basic and applied research in forensic entomology. *Annual Review of Entomology, 56*, 401–421.

Toom, V. (2018). *Cross-border exchange and comparison of forensic DNA data in the context of the Prüm Decision. Civil liberties, justice and home affairs*. http://www.dnaresource.com/documents/2008INTERPOLGLOBALDNASURVEYREPORTV2.pdf.

Van Dijk, E. L., Auger, H., Jaszczyszyn, Y., & Thermes, C. (2014). *Trends in Genetics, 30*, 418–426.

Wendt, F. R., King, J. L., Novroski, N. M. M., Churchill, J. D., Ng, J., Oldt, R. F., McCulloh, K. L., Weise, J. A., Smith, D. G., Kanthaswamy, S., & Budowle, B. (2017). Flanking region variation of ForenSeq DNA signature Prep kit STR and SNP loci in yavapai native Americans. *Forensic Science and International Genetics, 28*, 146–154.

Williams, R., & Wienroth, M. (2017). Social and ethical aspects of forensic genetics: A critical review. *Forensic Science Review, 29*(2), 145–169. https://www.nytimes.com/2021/05/31/science/dna-police-laws.

CHAPTER 25

Validation of NGS for casework at forensic DNA laboratories

Deborah Soares Bispo Santos Silva
Department of Chemistry, Hofstra University, Hempstead, NY, United States

Introduction

Next-generation sequencing (NGS), also known as massively parallel sequencing (MPS), is a high-throughput DNA sequencing technology capable of processing millions of DNA strands in parallel, yielding a higher resolution than capillary electrophoresis (CE), the current gold standard method used for human identification. NGS technology has advanced rapidly, and it is causing a huge impact in different areas, including in the forensic field. NGS is a robust method that allows sequencing of all types of nucleic acids and simultaneous analysis of hundreds of genetic markers, such as short tandem repeats (STRs) and single nucleotide polymorphisms (SNPs), and it provides more results in faster time (Ballard et al., 2020). The advent of NGS technology presents forensic DNA laboratories with the opportunity to solve challenging cases involving degraded samples, complex mixtures, or cases with a limited amount of DNA. Also, forensic scientists are able to simultaneously obtain information on human identification, ancestral origin, lineage and kinship, and phenotypic traits (Bruijns et al., 2018). NGS is revolutionizing DNA analysis and is becoming a powerful tool for the forensic community, providing meaningful information and resolving challenges in forensic DNA casework.

For over 20 years, forensic DNA testing has been based on CE technology, which presents limitations when dealing with difficult samples. Often, limited amounts or degraded samples pose challenges for CE-based testing, which may lead to inconclusive results. NGS offers forensic laboratories a new technology to add to their list of available testing methods. The ability to obtain additional information from smaller amounts of DNA, as well as from degraded and mixed DNA samples, positions NGS as an indispensable tool for forensic DNA analysis.

But despite the advantages presented by this new technology, the widespread adoption of NGS in forensic analysis has been slow. When compared to other fields, forensic science has been slow in fully embracing and incorporating NGS. This is mainly due to the new workflow, which differs from CE and requires specialized training and validation of the methods for casework. The introduction of NGS technology in a forensic laboratory requires changes in sample processing, instrumentation, data analysis, and interpretation. Generally, forensic laboratories have limited resources

Next Generation Sequencing (NGS) Technology in DNA Analysis
ISBN 978-0-323-99144-5, https://doi.org/10.1016/B978-0-323-99144-5.00004-4
© 2024 Elsevier Inc.
All rights reserved.

available, and the combination of a large upfront cost of instrumentation with increased time and effort for validation and analyst training may be creating a roadblock for the swift adoption of NGS. Alonso and coauthors (Alonso et al., 2017) published the results of a survey conducted by the DNASeqEx consortium in collaboration with the European Network of Forensic Science Institutes (ENFSI) DNA working group that collected information on NGS from 33 European laboratories from 25 countries. Of these laboratories, 52% had purchased at least one NGS instrument by 2017. The ones that had not purchased any instruments listed the following reasons for not doing so: general costs of NGS, lack of funding, and hesitation to use NGS until applications were more developed. For those that purchased an instrument, implementation challenges were reported, such as a lack of consistent nomenclature and reporting standards, a lack of compatibility with existing national DNA database infrastructure, a lack of population data for statistical purposes, and a lack of an adequate legislative framework (Alonso et al., 2017).

To accelerate the adoption of NGS technology by forensic laboratories, companies manufacturing kits and instrumentation have worked on lowering costs. Also, there are now more peer-reviewed publications on NGS-STR population data (Casals et al., 2017; Gettings et al., 2018; Hussing et al., 2019; Moura-Neto et al., 2021; Novroski et al., 2016; Phillips, Devesse, et al., 2018; Silva et al., 2018, 2020; van der Gaag et al., 2016; Wu et al., 2019), NGS testing, and validation work (Brandhagen et al., 2020; Buchard et al., 2016; Cihlar et al., 2020; Faccinetto et al., 2021; Hollard et al., 2019; Jäger et al., 2017; Köcher et al., 2018; Senst et al., 2022; Silvery et al., 2020). In 2016, the International Society for Forensic Genetics (ISFG) addressed the issue of the compatibility between sequence-based STR data and CE-generated genotypes and the need for a consistent and standardized nomenclature framework for NGS-STR sequences (Parson et al., 2016). Two years later, the DNA Commission of the ISFG revised the minimal STR sequence nomenclature requirements and published a revised sequence template file and nomenclature format (Phillips, Gettings, et al., 2018). More recently, the STRAND (STR: Align, Name, Define) ISFG working group held a meeting to address the issue of a lack of a universally accepted nomenclature system. Attendees discussed ideas with the goal of progressing on a nomenclature scheme, and the summary of the meeting report was published in September 2019 (Gettings et al., 2019). The ultimate goal of this working group is the release of an official ISFG recommendation on the requirements for labeling STR sequences. Other developments to encourage laboratories to adopt and validate NGS for DNA casework were the publication of the revised Scientific Working Group on DNA Analysis Methods (SWGDAM) validation guidelines for DNA analysis methods (Scientific Working Group for DNA Analysis Methods (SWGDAM), 2016) and the addendum to "SWGDAM interpretation guidelines for sutosomal STR typing by forensic DNA testing laboratories" to Address NGS (Scientific Working Group for DNA Analysis Methods (SWGDAM), 2019a).

NGS kits and platforms for generating DNA profiles

Law enforcement agencies around the world have access to DNA profiles stored in national and international DNA databases, which usually contain information on STR loci and mitochondrial DNA (mtDNA). In the US, for example, the National DNA Index System (NDIS) accepts autosomal and Y chromosome STR profiles and mtDNA control region data (Federal Bureau of Investigation (FBI), n.d.). Currently, NGS kits can generate DNA profiles based on data for generally accepted DNA loci and also on locations that are not usually encountered in DNA databases, such as X-STRs and SNPs.

The ForenSeq DNA Signature Prep kit (Verogen, CA) was the first NGS kit commercially available for forensic purposes and also the first NGS-STR kit approved for upload to NDIS. This kit is part of an integrated workflow that includes the MiSeq FGx sequencing instrument platform and the ForenSeq Universal Analysis Software. With this system, investigators can generate DNA profiles for up to 96 DNA samples on a single sequencing run. The results include information on autosomal, X-, and Y-STRs, as well as SNPs for the purposes of identification and phenotypic and biogeographical ancestry information (Verogen, n.d.a). Other ForenSeq kits are also available for forensic purposes and are also used with the MiSeq FGx sequencing instrument platform and the ForenSeq Universal Analysis Software. These kits include the ForenSeq mtDNA Whole Genome Kit, for sequencing the whole mitochondrial genome (Verogen, n.d.e); the ForenSeq mtDNA Control Region Kit, for sequencing the control region of the mitochondrial genome (Verogen, n.d.d); the ForenSeq Kintelligence Kit, for forensic genetic genealogy purposes (Verogen, n.d.b), and the ForenSeq MainstAY Kit, which targets core autosomal and Y-STRs (Verogen, n.d.c).

The PowerSeq CRM Nested System (Promega, WI) was the first NGS-mtDNA kit approved for upload to NDIS. This system generates information on mitochondrial HVI, HVII, and HVIII control regions, and it is part of a workflow compatible with MiSeq technologies (Illumina, CA) and the GeneMarker HTS Software (SoftGenetics, LLC) (Promega, n.d.b). Another PowerSeq product is the PowerSeq 46GY System, a non-NDIS-approved kit as of yet that targets autosomal and Y-STR loci (Promega, n.d.a).

The Applied Biosystems Precision ID mtDNA Whole Genome Panel (Thermo Fisher, MA) is also NDIS-approved. However, only the control region results are allowed to be uploaded to the database. This NGS panel is part of a system that includes the Ion GeneStudio S5 System, the Ion Chef System, and the Converge Software with the NGS data analysis module. The Precision ID mtDNA Control Region Panel, which targets hypervariable (HV) regions I, II, and III, is also available. Other panels include the Precision ID Ancestry Panel containing SNPs that can provide biogeographic ancestry information; the Precision ID Identity Panel, which includes 34 upper Y-clade SNPs and 90 autosomal SNPs; and the Precision ID Global Filer

NGS STR Panel containing autosomal STRs and Y markers (Thermo Fisher Scientific Applied Biosystems, n.d.). It is worth noting that these last four cited kits are not NDIS-approved at this moment.

Milestones for NGS in forensic casework

As forensic laboratories get more exposed to NGS and become more acquaintance with its advantages, some important milestones are being set by different laboratories around the world that are implementing parallel sequencing in their casework. France, for example, started to analyze DNA samples with NGS technology for mixture deconvolution, analysis of complex kinship, and obtention of the genetic profile from highly degraded DNA. In 2018, during the GCC Forensics Exhibition and Conference, Dr. François-Xavier Laurent presented how the *Institut National de Police Scientifique* (INPS), in France, implemented a fully automated workflow using NGS technology (Laurent, 2018). Dr. Laurent, at that time, was the head of research and development in forensic genetics at INPS, which is the biggest forensic institute in France and comprises a network of five accredited forensic laboratories. Dr. Laurent presented a 2011 cold case where a conclusive result could not be obtained with the use of CE results alone. The ForenSeq Prep Kit and the MiSeq FGx system (Verogen) were then used, and additional autosomal STR data was recovered for two low-level minor male contributors. With this case, in February 2018, France became the first country to upload an NGS profile to a national DNA database.

The use of NGS technology was crucial to solving a sexual assault case in The Netherlands in 2019. This was the first criminal conviction based on DNA profiles obtained with an NGS system (Verogen, n.d.f). In 2015, a 28-year-old woman was sexually assaulted. Evidence from the crime was submitted for analysis, and CE was used to generate DNA profiles. A hit was found in the Dutch convicted criminal database, and a suspect was arrested. The defense lawyer challenged the interpretation of the CE results, and the suspect was later released by the judge. The analyzed sample contained mixed DNA, and based on only CE results, it was not possible to determine if some alleles were stutters from the victim's profile or if they belonged to a minor contributor. After an appeal by the prosecution, the use of Verogen's NGS technology was allowed, and it was performed by the Forensic Laboratory for DNA Research (FLDO), an independent forensic DNA laboratory at Leiden University Medical Center. This time, results were clear due to a great advantage of NGS: investigators not only had information on the size of each allele, but they could also analyze their sequences. They were able to identify more of the minor alleles, and after the use of likelihood ratio statistics, a conclusive result was obtained. After this first court case, FLDO, which was the first forensic laboratory to receive an ISO-17025 accreditation for NGS technology, has used this method in more than 30 cases. Investigators were able to generate profiles for some cold cases, and for other ones, NGS-generated profiles allowed potential suspects to be excluded.

Recently, Finaughty and coauthors (Finaughty et al., 2020) reported a case from 2018 where two sets of human remains (A and B) were discovered within 2 months of each other off the coast of Cape Town, South Africa. Both sets were unidentifiable due to marine decomposition, and preliminary analysis by the forensic pathologist suggested that the remains of set A belonged to a female individual. DNA analysis was then performed to determine the biological sex of the remains and if sets A and B belonged to the same individual. DNA profiles were obtained using both CE and NGS analyses, and according to the results, the remains belonged to the same male individual and not a female, as previously suggested. Although the results obtained using the conventional CE method and NGS were concordant, sequencing analysis provided higher random match probability values. According to the authors, this was the first forensic case in Africa where NGS and the MiSeq FGx system were successfully used and assessed. Also, this was the first report of successful DNA profiling with an NGS system using soft tissue lysates from marine decomposition.

Back in 2019a, Verogen announced a partnership with Cellmark Forensic Services, a provider of forensic DNA services in the United Kingdom (UK), to establish Cellmark as a UK center of excellence for forensic NGS (Verogen, 2019a). As part of the new collaboration, Verogen would supply Cellmark with its ForenSeq DNA Signature Prep Kit. The goal of this collaboration was to provide the capabilities of NGS to police investigators, such as simultaneous analysis of different DNA markers to generate data on identification, appearance, and biogeographical ancestry, and the potential to recover more information from degraded and mixed samples. Two years later, in September 2021, Cellmark became the first forensic laboratory to receive ISO17025 accreditation from the UK Accreditation Service for forensic sequencing services (Cellmark Forensic Services, n.d.).

In the United States, important milestones were reached when the FBI approved forensic NGS systems to be used by laboratories for the generation of DNA profiles for NDIS. In May 2019, the MiSeq FGx system (Verogen) became the first NGS NDIS-approved system for STR profiling. This was achieved by the submission of internal validation results to the FBI by four different laboratories: the Washington, D.C., Department of Forensic Sciences (DFS), the Armed Forces DNA Identification Laboratory (AFDIL), the University of North Texas Health Science Center (UNTHSC), and the Ohio Bureau of Criminal Investigation (OBCI) (Verogen, 2019b). Also in May 2019, the PowerSeq CRM Nested System (Promega), which is used for generating mtDNA profiles, was approved for submission to the US NDIS Combined DNA Index System (CODIS) database (Promega, n.d.b). In July 2019, a second mtDNA NGS system became NDIS-approved, Thermo Fisher Scientific's Applied Biosystems Precision ID System (Thermo Fisher Scientific, n.d.). It is important to mention that the mtDNA profiles generated by both mtDNA kits can only be used for missing person-related searches at NDIS.

Validation of NGS-based DNA analysis methods

The use of CE-based forensic DNA typing is well established in forensic laboratories and widely accepted in court. Before the implementation of NGS technology into forensic practice and prior to its use in casework, its performance needs to be validated. Any new assay or method introduced in a forensic laboratory must be vigorously tested to ensure that the procedure works properly and that the results obtained are in accordance with the intended purpose (Butler, 2012). Through validation studies, procedural limitations are defined, and the method is shown to be robust, repeatable, and reliable. Once the process of validation is completed, standard operating procedures (SOP) and interpretation guidelines are established, and the new method can be used routinely in the laboratory workflow.

Method validation is required by the ISO/IEC 17025 accreditation, which establishes the requirements for laboratory competence to carry out tests and/or calibrations. However, ISO/IEC 17025 only determines a general standard, and experts in the forensic field should give more detailed recommendations (European Network of Forensic Science Institutes (ENFSI), 2010).

On July 1st, 2020, the FBI published updated quality assurance standards (QAS) for forensic DNA testing laboratories. These standards must be followed by laboratories performing forensic DNA testing or utilizing CODIS to ensure the quality and integrity of the data generated by the laboratory (Federal Bureau of Investigation (FBI), 2020).

The promising implementation of NGS into the forensic casework routine led to the formulation of additional validation guidelines by SWGDAM. SWGDAM is constituted of a group of invited guests, representing international organizations or laboratories, academia, and accrediting agencies. SWGDAM issues guidelines for implementation by crime labs and also provides recommendations to the FBI Director on quality assurance standards for forensic DNA analysis (Scientific Working Group for DNA Analysis Methods (SWGDAM), n.d.). It is important to point out that SWGDAM issues guidelines, not minimum standards. In the case of a conflict between the FBI's QAS and recommendations given by SWGDAM, the QAS and the QAS Audit Documents have precedence over these guidelines (Scientific Working Group for DNA Analysis Methods (SWGDAM), 2016).

There are typically two types of validation: developmental and internal. Developmental validation involves testing new DNA methods for use on forensic samples to determine their conditions and limitations. This type of validation is usually done by commercial manufacturers and large laboratories (Butler, 2012). It can also be performed by smaller laboratories in the following situations: (a) if involved in the development and testing of a new method with a commercial company; (b) if involved in developing and testing a new in-house method; or (c) if a previously validated method is significantly modified or applied to a new situation and needs revalidation. Internal validation is done to verify if established methods and procedures will work as expected in one's

own laboratory (Butler, 2012). This type of validation is usually done by smaller local and state forensic DNA laboratories before implementing a previously validated (by developmental validation) method or procedure (Butler, 2012; National Institute of Standards and Technology (NIST), n.d.; Scientific Working Group for DNA Analysis Methods (SWGDAM), 2016).

National DNA Index System

The National DNA Index System, best known by its acronym NDIS, is a database of DNA profiles obtained and uploaded by federal, state, and local participating forensic laboratories in the United States. The DNA Identification Act of 1994 allowed the establishment of NDIS and provided the categories of information maintained in the database, as well as the requirements for quality assurance, privacy, and expungement that participating laboratories must follow (Federal Bureau of Investigation (FBI), 2021).

Currently, the DNA data accepted by NDIS include autosomal and Y- short tandem repeats (STRs), and mitochondrial DNA (mtDNA). Effective May 1, 2019, the DNA data submitted to NDIS must be generated in accordance with SWGDAM interpretation and validation guidelines and must follow guidelines outlined in the Addendum to "SWGDAM interpretation guidelines for autosomal STR typing by forensic DNA testing laboratories" to address the NGS (Scientific Working Group for DNA Analysis Methods (SWGDAM), 2019a), the SWGDAM interpretation guidelines for mitochondrial DNA analysis by forensic DNA testing laboratories (Scientific Working Group for DNA Analysis Methods (SWGDAM), 2019b), and the SWGDAM validation guidelines for DNA analysis methods (Scientific Working Group for DNA Analysis Methods (SWGDAM), 2016).

The current NGS PCR kits accepted for use at NDIS are:

- Verogen ForenSeq DNA Signature Prep Kit
- Promega PowerSeq CRM Nested System
- Thermo Fisher Scientific Precision ID mtDNA Whole Genome Panel—only the mtDNA control region data is approved for NDIS
- Verogen ForenSeq MainstAY Kit

FBI's quality assurance standards

The validation studies described in the latest version in effect in July 2020 can be applied to the implementation of NGS systems (Federal Bureau of Investigation (FBI), 2020). Crime labs must validate all new NGS workflows, including modified portions of previously validated and approved systems.

According to QAS Standard 8.2, developmental validation requires (when applicable) the following set of studies to be performed:

- Characterization of the genetic marker: it is important to understand and document locus characteristics of the genetic markers used, such as inheritance, mapping, method of detection, and polymorphism(s).

- Species specificity: nonhuman studies are conducted to evaluate if nonhuman DNA can interfere with the results of the analysis of forensic samples. The assay is not necessarily invalidated if genetic information is recovered from these nonhuman samples, but the data obtained should be used to define the limits of the assay.
- Sensitivity studies: it is important to determine the maximum and minimum DNA quantities necessary to produce reliable results.
- Stability studies: known samples are deposited on various substrates and are also put under different types of insults, such as environmental and chemical, to evaluate the effects of such conditions on the obtained results.
- Case-type samples: mock or real forensic samples (usually from closed cases) are used to demonstrate reliable results with casework samples.
- Database-type samples: for this type of study, the sample types and/or sample substrates tested are those routinely submitted to the database laboratory.
- Population studies: it is important to determine allele frequencies for major population groups to use in the calculation of random match probability. Tests for independence expectations, such as Hardy-Weinberg equilibrium and linkage equilibrium, must be performed.
- Mixture studies: it is important to determine the limit of detection when working with mixed specimens, especially in the detection of major and minor components in samples with more than one contributor. Tested mixed DNA samples must be representative of those typically encountered by the testing laboratory.
- Precision and accuracy studies: repeatability and/or reproducibility should be addressed. This type of study will demonstrate how consistent the results obtained are and how close they match the actual values.
- PCR-based studies: this type of study will demonstrate the reaction conditions needed to obtain reliable results and often include studies on thermal cycling conditions (including primer concentration, DNA polymerase, and other reagents necessary to the reaction), assessment of differential and preferential amplification, effects of multiplexing, assessment of appropriate controls, and product detection studies.

All developmental validation studies must be documented, and publication(s) supporting the underlying scientific principle(s) of a method must be available. This way, others can understand and assess the validation work that was performed. Also, peer-reviewed publications are encouraged due to their importance in the legal system. These publications show that the methodology or technology is generally accepted in the scientific field, and courts can decide whether or not the results obtained are admissible.

Standard 8.3 describes the internal validation, which requires, as applicable, the following set of studies to be performed:

- Known and nonprobative evidence samples or mock evidence samples: Well-characterized known samples, such as NIST Standard Reference Materials, should be used to test the new method or technology to demonstrate that it is working as

expected. Then, the following tests should be done using real case samples or mock case-type samples.

- Known database-type samples: known samples, available database samples, or mock samples (these should reflect the type of samples encountered in databasing) should be tested and compared to previous results where possible for concordance.
- Precision and accuracy
- Sensitivity and stochastic studies: stochastic effects are common in PCR-based methods when working with low amounts of DNA. Therefore, it is important to test the limits of sensitivity to understand how it can affect your results and ensure accurate data. However, understanding just the lower limit is not enough. Testing the upper limitations is crucial to know the optimal range of the method.
- Mixture studies
- Contamination assessment studies: minimizing and assessing contamination is essential when working with forensic samples. Negative controls should always be included during extraction and amplification procedures to evaluate the detection of exogenous DNA. The results of this type of study should be used for the development of quality control procedures and interpretation guidelines.

SWGDAM validation guidelines for DNA analysis methods

The document providing guidelines for the validation of DNA analysis methods was revised in November 2016 to address NGS technologies, and it was approved and posted in December 2016 (Scientific Working Group for DNA Analysis Methods (SWGDAM), 2016). It is important to know that this new document supersedes the SWGDAM revised validation guidelines published in 2012.

Laboratories will first determine the type of validation to conduct and then, which studies to perform, depending on the methodology and its application, and which number of samples will be needed for each study of the validation plan.

Developmental validation process

Section 3 of the validation guidelines describes what shall be included, when applicable, in the developmental validation process. Most studies are concordant with the FBI's QAS, except for one that addresses NGS technology specifically. The following are the studies listed in the developmental validation guidelines:

- Characterization of the genetic marker (subsection 3.1)
- Species specificity (subsection 3.2)
- Sensitivity studies (subsection 3.3)
- Stability studies (subsection 3.4)
- Precision and accuracy (subsection 3.5)
- Case-type samples (subsection 3.6)
- Population studies (subsection 3.7)

- Mixture studies (subsection 3.8)
- PCR-based studies (subsection 3.9)
- NGS-specific studies (subsection 3.10)

Developmental validation process—subsection 3.10

According to the updated guidelines, whenever a laboratory is validating an NGS method, they should:

- Assess the effects of barcoding/indexing samples and subsequent bioinformatic sample separation. When appropriate, the effects of novel barcodes/indices that have not otherwise undergone developmental validation should be included (subsection 3.10.1).
- Address the limit of detection as it relates to both starting DNA input as well as the extent of varying the quantity (extent of sample multiplexing in sequencing) and/or quality of libraries pooled in the sequencing reaction (subsection 3.10.2, sensitivity studies).
- Assess NGS instrumentation for the possibility of signal cross-talk during sequencing and sample carryover between runs (subsection 3.10.3).

In NGS methods, multiple libraries are pooled and sequenced together in a single run. This is possible because of the use of barcodes/indices, which are unique index sequences added to each library to distinguish between them during data analysis. However, sequencing or demultiplexing errors can occur, and one way to evaluate this is by using bioinformatic negatives, which consist of using barcode/indices combinations different than the ones used for samples in the run. Also, runs might be performed in sequence, making it important to use the bioinformatic negatives to detect any carryover from one run to the other. For example, if a lab is doing two sequencing runs and the second one is right after the first one is finished, the barcode combinations used should be different for each run. This way, when analyzing the second run, bioinformatic negatives can be included, and the barcode combinations used in the first run will be used in the analysis of the second run. If no carryover or contamination is present or if there are no demultiplexing errors, the reads assigned to the bioinformatic negatives should be very low and well below the assigned read threshold.

Also, samples available for analysis vary in quantity and quality. For this reason, during the validation process, it is important to sequence a low number of samples in the same run and include samples with higher DNA quantity and quality, such as positive controls, samples with low DNA input and/or lower quality, and also negative controls. Then the results from this run can be compared with those from a run consistent with larger library pools. This type of study will verify the possibility of sample crosstalk and determine if the size and quality of the sample pool will influence the data recovered.

Internal validation process

Section 4 of the validation guidelines describes what shall be included, when applicable, in the internal validation process. Again, most studies are concordant with the FBI's QAS. Developmental validation studies conducted within the same laboratory may satisfy some elements of the internal validation process.

The following are the studies listed in the internal validation guidelines:

- Known and nonprobative evidence samples or mock evidence samples (subsection 4.1)
- Sensitivity and stochastic studies (subsection 4.2)
- Precision and accuracy (subsection 4.3)
- Mixture studies (subsection 4.4)
- Contamination assessment (subsection 4.5)
- NGS-specific studies (subsection 4.6): 4.6.1 Sensitivity studies should address the limit of detection as it relates to both starting DNA input as well as the extent of varying the quantity (extent of sample multiplexing in sequencing) and/or quality of libraries pooled in the sequencing reaction.

Addendum to "SWGDAM interpretation guidelines for autosomal STR typing by forensic DNA testing laboratories" to address next-generation sequencing

The SWGDAM interpretation guidelines document for autosomal STR typing, also mentioned in this chapter as the parent document, acts as a guide for laboratories in the interpretation of DNA typing results from STRs, and its latest version was approved on January 12, 2017, and revised on July 13, 2021 (Scientific Working Group for DNA Analysis Methods (SWGDAM), 2021). On April 23, 2019, an addendum providing guidelines for the interpretation of autosomal STR data developed via NGS was approved by the SWGDAM membership (Scientific Working Group for DNA Analysis Methods (SWGDAM), 2019a). The publication of this addendum is extremely important to assist forensic laboratories in the process of validation and implementation of NGS technology. Important guidelines are given on sequence-based allele designation, analytical and stochastic thresholds, mixture interpretation, and statistical analysis based on sequence data. According to SWGDAM, these guidelines are for laboratories that plan on using binary approaches to interpret sequence-based STR data. It is important to remember that sequence-based STR data can be converted to fragment length-based STR alleles, making it backward compatible with length-based data generated by conventional methods such as CE.

Forensic laboratories are used to interpreting fragment length-based data and are familiar with terms such as peak, peak height ratio (PHR), and relative fluorescence units (RFUs). When working with NGS and interpreting data generated by this type of tech-

nology, laboratories should get familiar with the new terms. To facilitate the data interpretation, a simplified way to understand the new terminologies is to make the following correlations:

- Peak = Signal
- PHR = Allele count ratio (ACR)
- RFU = Read count

Section 1 of the addendum goes over the interpretation of DNA typing results. Unlike CE-generated data, where an electropherogram is created based on fluorescence intensity data and peaks are labeled by the software using allelic ladders as a reference, NGS-generated data need a software program to demultiplex the samples in the sequenced library, detect the sequence of nucleotides in each generated strand, and then convert the final information into sequence reads. For each locus, the reads are combined, and total sequence read counts are reported along with other allele-specific information, including each allele designation with its sequence, length, nomenclature, and read count. Similar to CE results, NGS final results need to be verified by a DNA analyst to ensure the accuracy of the data generated by the software. First, it is verified that the run performance and the quality of the data generated meet the established performance and quality metrics, which may include the amount of raw data and the quality of the generated libraries. Second, the reagent blank(s), positive control(s), and negative control(s) are verified to check if they present the correct genotyping results. Only then the DNA analyst can analyze each sample in the sequence run. Also, if using multiple kits and/or platforms, the DNA analyst must check if the data generated is concordant between the overlapping loci of the different kits.

The following are the main subsections that apply to the interpretation of NGS-generated autosomal data. Some subsections are only present in the parent document, and there are no NGS-specific guidelines in the addendum.

Subsection 1.1. Analytical threshold (AT) is important to allele detection and should be set to distinguish between a real signal and background noise. ATs must be defined based on internally derived empirical data. For NGS methods, AT represents the minimum read count at or above which all detected signals are reliable and should be interpreted. There are multiple ways to set ATs, and they can be a fixed read count value or a percentage. As referenced in the addendum document, this last type of AT is known as a percentage-based threshold, and it can be calculated by dividing the allele read count by the total locus read count. When choosing ATs, laboratories must take into consideration that nonreproducible noise may be detected above the set AT. To minimize this from happening, some labs might increase the AT value. However, it is important to take into consideration that raising the AT too much may result in data loss. The other extreme is also possible, and careful consideration must be taken not to set too low an AT value and run the risk of having erroneous information added to the results and/or having an overinterpretation of the data. Defining background noise can be challenging, and

background noise levels may vary by sample and by locus within a sample. For this reason, different ATs can be set for different loci instead of using the same AT value for all loci analyzed.

Subsection 1.2. The laboratory must evaluate the sequencing run based on previously established quality metrics. The lab can choose to evaluate the sequencing run based on results from positive controls and sequencing standards or based on run metrics defined during the validation process. The run metrics can include loading density, cluster density, clusters passing the filter, phasing and prephasing values (percentage of molecules within a cluster that fall one base pair behind or ahead in each cycle), total reads per sample, total reads per run, forward/reverse strand balance, Q-scores (estimate of the probability of that base being called wrongly), etc. These metrics will only provide an overview of run performance, not the performance of each sample in the run.

Subsection 1.3. Controls must be included in each run, and laboratories must establish criteria on how to evaluate them. Controls may include reagent blanks and positive and negative controls. Additional controls may also be included for troubleshooting and quality control purposes.

Subsection 1.4. Locus designations and locus assignment for alleles must be done based on established criteria by the lab. Positive controls can be used for the verification of correct designations. In this case, the lab needs to confirm that the software used for analysis has been validated for this purpose. Locus designations must include the range of the reported sequences, which should be based on a relevant human genome reference sequence (e.g., hg 19, GRCh38, etc.).

Subsection 1.5. Allele designation must be done using length-based and/or sequence-based data. SWGDAM has yet to provide additional guidance on STR sequence nomenclature but suggests that allele designation, as a numerical value or as a sequence, be done in accordance with the guidance provided by the ISFG (Parson et al., 2016; Phillips, Gettings, et al., 2018). When the allele designation is sequence-based, there can be a variation in the sequence range reported depending on the NGS assay and/or analytical software used. The reported sequence may include the repeat region only or may also include additional information on the flaking region. When the allele designation is length-based, it should be done according to the fragment size and not just based on the number of repeat units in the core region. This method of designation should maximize concordance with CE methods, and the developmental validation of the analytical software is crucial to ensure this concordance between NGS and CE platforms. The sequencing of STRs has enabled the identification of isoalleles, which are alleles with the same size (length) but different sequences (Table 25.1). For this reason, laboratories must have established criteria to report isoalleles if sequence data is used for interpretation. However, even if the sequence data is mostly used, laboratories should search databases using the length data due to backward compatibility. Discrepancies between NGS-developed length-based and sequence-based alleles can arise due to flanking region

Table 25.1 Representation of isoalleles in locus D12S391. Notice that all alleles have the same length but different sequences.

Locus	Chromosome	Allele (length)	Bracketed repeat region
D12S391	12	22	[AGAT]11 [AGAC]11
D12S391	12	22	[AGAT]12 [AGAC]10
D12S391	12	22	[AGAT]12 [AGAC]9 AGAT
D12S391	12	22	[AGAT]13 [AGAC]9

insertions or deletions, for example. In these situations, laboratories should have an established policy for addressing and documenting these discrepancies, including, if possible, a reason for the nonconcordance. Also, there must be established guidelines for the designation of alleles containing an incomplete repeat motif or any reported sequence resulting in other than a whole-number allele. And finally, previously unobserved sequences and/or unusual motifs may be encountered with the sequencing of autosomal STRs. In the event of this type of situation, the following steps are suggested: Consider the data quality metrics, such as the ones described in subsection 1.2, before recognizing the new sequence or motif; perform a literature search to confirm that the encountered sequence or motif has not been previously published; resequence if the sequence or motif is not found in the literature; and consider if the unusual sequence has been properly designated by the software (length-based and sequence-based designations).

Subsection 1.6. Nonallelic signals may be present in the final data and must be correctly identified. These signals may be PCR products, such as a stutter, or analytical artifacts. Validation studies are important to aid in the identification of nonallelic signals. Specific empirical data for each NGS assay and detection system must be used to establish interpretation guidelines. Noise signals can be identified based on the sequence of the artifact and/or its reproducibility. A stutter signal can be identified if its sequence differs in length, or the number of repeat motifs, when compared to the parent allele, and if it exhibits a lower read count in relation to the authentic allele. An example can be found in Table 25.2. Well-defined thresholds based on validation data will aid in the identification of real alleles versus nonallelic signals and in determining what to expect with regard to the proportion of noise/stutter relative to a real allele. Because the proportion levels can vary with the increase or decrease of read counts, ATs that are locus-specific might be needed, as previously discussed in subsection 1.1.

Table 25.2 Representation of an authentic allele and its sequence, length, and read count, and a stutter sequence at 8.9% of the parent allele signal 2.

Locus	Chromosome	Allele (length)	Read count	Bracketed repeat region
D12S391	12	22	35	[AGAT]12 [AGAC]10
D12S391	12	23	395	[AGAT]13 [AGAC]10

Subsection 1.7. Stochastic effects are common when amplifying low-level DNA samples. When working with CE methods, different peak heights may appear for a heterozygous locus, or an allele dropout can occur. For this reason, a specific peak height value is defined as a stochastic threshold. For NGS methods, instead of differences in peak heights, there can be differences in read counts, with the minor allele presenting a lower number of sequencing reads, leading to a heterozygous read count ratio imbalance or an allele dropout. A fixed read count stochastic threshold and a percentage-based threshold should be established and can be applied to all loci analyzed or be locus-specific. Stochastic thresholds must be based on empirical data and be specific to the amplification and detection systems used. Again, this shows the importance of validation studies and how the data obtained can aid in the determination of thresholds.

Subsection 1.9. The estimation of the number of contributors to a DNA profile can be done based on the number of alleles present at all loci. If one or two alleles are present in each locus, then it's usually assumed that there is only one contributor. If three or more alleles are present at one or more loci, then this can be indicative that there are multiple contributors. When working with a CE-generated profile, PHRs may also be analyzed and compared to the empirically determined heterozygous peak height ratio expectation. For NGS-generated profiles, PHRs are not applicable, but information on isoalleles can be used, and the number of alleles with different sequences but equal lengths can be indicative of the number of contributors.

Section 2 of the addendum goes over the overview and strategies for mixture interpretation. The parent document goes over the guidelines for STR mixture interpretation based on CE data. If a laboratory is working with length-based NGS data, then the same guidelines will apply. The addendum focuses on the differences in interpreting mixtures when sequence-based NGS data are used. The following are the main subsections that apply to the evaluation of mixture profiles at the sequence level. Some subsections are only present in the parent document, and there are no NGS-specific guidelines in the addendum.

Subsection 2.1. All criteria used for mixture interpretation shall be supported by the data and shall be defined and documented. This includes all assumptions that are made based on information obtained from isoalleles. When working with sequence data, the detection of isoalleles can aid in the interpretation of the data and in distinguishing alleles from minor and major contributors.

Subsection 2.2. When dealing with two-person mixtures, pair-wise comparison of all potential genotypic combinations should consider sequence data and allele count ratios (also known as ACR, which is analogous to CE-based peak height ratio) of all sequence combinations, including isoalleles.

Subsection 2.3. When dealing with greater than two-person mixtures, the interpretation of data must be based on results obtained from sequence-based mixture studies, including known contributors with isoalleles.

Subsection 2.4. Again, the presence of isoalleles will help when interpreting a mixed sample. When dealing with mixtures with a single major contributor and one or more minor contributors, the sequence data may differentiate isoalleles, and this may allow the determination of major and minor sequence alleles that are the same length.

Subsection 2.6. Sequence information can be helpful when interpreting potential stutters in a mixed sample because it can allow the differentiation between stutter and same-length minor contributor alleles. And once more, validation studies are important to provide information on sequence-based stutter expectations.

Section 3 is the final section of the document and goes over the statistical analysis done after the interpretation of the data is finalized. The parent document goes over the guidelines for CE-based comparisons of reference profiles to evidence profiles. However, if a laboratory is working with length-based NGS data, then the same guidelines will apply. The addendum focuses on the differences encountered when comparing profiles at the sequence level. There is only one subsection that presents NGS-specific guidelines. The other subsections are only present in the parent document, and there are no NGS-specific guidelines in the addendum.

Subsection 3.2. Special attention must be paid to the reported range when comparisons of reference profiles to evidence profiles are done using sequence data. For this type of comparison, sequence-based allele frequency data shall be used, and the sequence range of the allele frequency data should be the same as the reported ones for the laboratory data. However, if the reported range for the frequency data is larger, then the laboratory can first truncate the published frequency data set to match their reported range. If the reporting range for the frequency data is smaller, then the laboratory can truncate the reported range for their allele sequence data to match the frequency data. For statistical calculations of sequence-based single-source genotypes, isoalleles are considered heterozygote genotypes, and the formula is 2pq. However, if laboratories are performing length-based analysis, then they should have policies for interpretation and statistical analysis when isoalleles are encountered to avoid concealing potentially exculpatory information.

And lastly, there is no current guidance regarding theta values for sequence-based data; therefore, the existing NRC II guidance should be followed (NRC II 4.4a, where typically $\theta = 0.01$ for most U.S. groups or 0.03 for some isolated populations).

SWGDAM interpretation guidelines for mitochondrial DNA analysis by forensic DNA testing laboratories

The SWGDAM interpretation guidelines document for mtDNA analysis was revised to address NGS, and it was approved on April 23, 2019 (Scientific Working Group for DNA Analysis Methods (SWGDAM), 2019b). This document supersedes the SWGDAM interpretation guidelines for mtDNA analysis by forensic DNA testing laboratories, published in 2013. Laboratories should develop and implement interpretation guidelines based on validation studies that follow the FBI QAS and the SWGDAM validation guidelines. The guidelines document is meant to assist in the interpretation of

results generated by Sanger or NGS technologies. mtDNA analysis is already sequence-based, and the use of NGS methods is well suited for this type of analysis. Because the final data format is very similar between Sanger-generated data and NGS-generated data, many sections of the interpretation guidelines document are applicable to both methods. For the purpose of this chapter, the focus will be on the subsections with NGS-specific guidelines.

Section 2 focuses on sequence analysis and nomenclature. In subsection 2.2.1, the criteria that must be established for the analysis of sequencing results are discussed. Laboratories should create guidelines to assign nucleotide base calls and to verify the quality of results, determining whether they are sufficient for interpretation purposes. All established criteria should be supported by validation studies. For NGS-generated data, individual laboratories should establish criteria for filtering and trimming reads based on validation data and/or described in the software literature. Read coverage, Q-scores, and strand bias should be considered when determining the quality and suitability of the sequence data. Sequencing errors can occur in a biased manner due to one-strand orientation. Laboratories can determine a strand bias score and filter out bases with extreme strand bias. Other metrics to consider for the designation of bases include variant frequency, count, and quality.

Subsection 2.2.4 discusses the interpretation of mixtures. It is not common practice to interpret Sanger-based mtDNA mixed sequences. Sanger sequencing of mtDNA allows the detection of mixtures but does not allow mixture deconvolution. For this reason, most forensic laboratories choose not to interpret the mixture results. However, it is a different scenario when working with NGS data due to its quantitative nature. When using NGS methods, laboratories have access to data on the number of reads, which allows the quantitative determination of the components in a mixed sample. For a laboratory to perform mtDNA mixture interpretation with NGS, validation studies must be conducted following SWGDAM validation guidelines. The limitations of the analysis must be fully characterized, and experiments to distinguish among mixture, heteroplasmy, damage, and noise must be performed. Once all studies are done, then the laboratory can establish a protocol for mtDNA mixture interpretation.

Section 2.3 focuses on the guidelines for sequence nomenclature. Independently of the method used to obtain mtDNA sequences, Sanger sequencing or NGS, the same SWGDAM nomenclature rules should be followed, and the revised Cambridge Reference Sequence (rCRS) described by Andrews and coauthors (Andrews et al., 1999) should be used as the consensus sequence.

The following are the SWGDAM nomenclature rules that analysts should follow to code variants from the rCRS:

- **Rule 1**—Maintain known patterns of polymorphisms (i.e., known phylogenetic alignments). Most violations of known patterns of polymorphism involve insertions and deletions. A phylogenetic alignment tool is available at empop.org.
- **Rule 2**—Use nomenclature with the least number of differences unless it violates known patterns of polymorphism.

- **Rule 3a**—Homopolymeric C-stretches in hypervariable region I (HVI): C-stretches in HV1 should be interpreted with a 16189C when the otherwise anchored T at position 16,189 is not present. Length variations in the short A-tract preceding 16,184 should be noted as transversions.
- **Rule 3b**—Homopolymeric C-stretches in hypervariable region II (HVII): C-stretches in HV2 should be interpreted with a 310C when the otherwise anchored T at position 310 is not present. C-stretches should be interpreted with a 311T when the anchored T at position 310 is followed by a second T.
- **Rule 4**—Maintain the AC repeat motif in the HVIII region from np 515–525.
- **Rule 5**—Prefer substitutions to insertions/deletions (indels).
- **Rule 6**—Prefer transitions to transversions unless this is in conflict with Rule 1.
- **Rule 7**—Place indels contiguously when possible.
- **Rule 8**—Place indels on the 3′ end of the light strand.
- **Rule 9**—The 3107 nucleotide should not be reported in sample data. As 3107 in the rCRS is simply a placeholder intended to maintain historical nomenclature (Andrews et al., 1999), differences from the rCRS (i.e., deletions) at this position are not biologically meaningful.

When working with NGS, as explained in subsection 2.3.4 (Length variants), the analysis software used should be taken into consideration when establishing the guidelines for the interpretation of C-stretches. Analysts should do a further review of these regions to ensure that variant calling metrics are met because different software programs used in NGS workflows may not align homopolymeric (and other) repeat regions uniformly or according to standard forensic practice, as stated in rules 1, 7, and 8 explained above. And once more, validation studies are important and should be done to characterize the performance of the analysis software and establish interpretation guidelines for C-stretch regions.

Peer-reviewed publications on validation studies of NGS systems

As previously mentioned, peer-reviewed publications of validation studies are not required but are encouraged. Many validation studies have been published to this date, and they have characterized and shown the limitations of established and NDIS-approved NGS workflows, as well as in-house-developed assays and those yet to be approved for upload in DNA databases. Here are some examples of published studies done to validate the three first NGS systems approved for upload to the NDIS.

In 2017, Illumina published the developmental validation of the MiSeq FGx Forensic Genomics System, comprised of the ForenSeq DNA Signature Prep Kit, MiSeq FGx Reagent Kit, MiSeq FGx instrument, and the ForenSeq Universal Analysis Software (Jäger et al., 2017). The ForenSeq DNA Signature Prep Kit is an assay comprised of two primer mixes: DNA Primer Mix A (DPMA) targets Amelogenin, 27 autosomal STRs, 24 Y-STRs, 7 X-STRs, and 94 identity-informative SNPs, and DNA Primer

Mix B (DPMB) targets each of the loci in DPMA, as well as biogeographic ancestry-informative SNPs and phenotypic-informative SNPs. Users may choose which primer mix they want to use for analysis. The studies were performed in accordance with the 2012 revised SWGDAM validation guidelines and encompassed species specificity, sensitivity, mixed samples, stability (inhibitor, degradation), accuracy, and precision. An important feature demonstrated in the studies performed was the backward compatibility of allele calling with current STR databases. Robustness was also assessed through sensitivity studies, DNA mixture testing, and stability analysis. Genomic DNA samples were partially degraded using mechanical shearing, and the effects of five known PCR inhibitors on amplification efficiency were characterized. For the mixture studies, three sets of two-person DNA mixtures were prepared in varied ratios. When analyzing data from all studies, quality indicators in the ForenSeq Universal Analysis Software assisted in evaluating all results. Quality indicators are an important feature of the software, and each crime laboratory should set thresholds, filters, and other analysis parameters based on data obtained during their internal validation studies to establish their own internal laboratory guidelines and policies. Repeatability and reproducibility studies were also performed and indicated 100% accuracy in allele calling when compared to CE data for STRs and >99.1% accuracy when compared to bead array typing for SNPs. Limitations were encountered when a higher number of high-quality samples were pooled together in a single sequencing run, and not all samples yielded full profiles. The authors suggested that depending on the level of profile completeness required in certain cases, individual laboratories should determine the ideal number of samples in a sequencing run. This type of result can be obtained when forensic laboratories perform internal validation studies. Other limitations include the imbalance encountered for the autosomal STR locus D22S1045; higher relative levels of intralocus imbalance for three SNP loci (rs10776839, rs7041158, and rs6955448); and higher stutter percentages for the rapidly mutating Y loci. Once again, individual laboratories should perform internal validation studies and establish interpretation guidelines for use in the analysis of their casework samples. Sensitivitiy studies showed that the input of DNA ranging from 62.5 pg to 1 ng produced complete, accurate, and reproducible genotypes. Overall, the results obtained in this study indicated that the MiSeq FGx System is robust, reliable, reproducible, and meets established forensic validation guidelines. Since the publication of this developmental validation work, other forensic laboratories have performed validation studies in order to develop their own interpretation guidelines and implement this NGS system in their forensic DNA analysis.

In 2021, Faccinetto and co-authors (Faccinetto et al., 2021) published internal validation studies done for the Precision ID mtDNA Whole Genome Panel (Thermo Fisher Scientific). This was a work done by the Department of Police Scientific Investigations (Parma, Italy) in collaboration with multiple universities in Italy. In this work, the typing of the mtDNA genome was done using the Precision ID mtDNA Whole Genome Panel v.2.2 (Thermo Fisher Scientific) and the Ion S5 system (Thermo Fisher Scientific). All

studies were performed in accordance with the SWGDAM validation guidelines for forensic DNA analysis methods and the ENFSI Recommended minimum criteria for the validation of various aspects of the DNA profiling process. The European guidelines were issued in 2010 and do not specifically address NGS workflows (European Network of Forensic Science Institutes (ENFSI), 2010). NIST Standard Reference Material 9947A and the 2800 M control DNA (Promega Corporation), as well as typical casework specimens, were used for concordance, repeatability, reproducibility, sensitivity, and heteroplasmy detection analyses. All results were compared with data generated by Sanger sequencing and another NGS sequencer in a different laboratory. With the results, the authors encountered strengths and limitations, mostly related to noise thresholds and heteroplasmy detection. To mitigate such issues, a filtering step for reads was applied to the analytical workflow to reduce the noise level, and the authors suggested bioinformatic solutions using a probabilistic approach based on the analysis of the distribution of variant frequencies to allow the estimation of probabilities for each point heteroplasmy (PHP) call. Overall, results from the validation studies showed that the Precision ID Whole mtDNA Genome Panel is accurate, reproducible, reliable, and sensitive, enabling the generation of useful full mtDNA sequences from challenging samples common in forensic casework and proving to be more sensitive than Sanger sequencing.

An important publication on validation studies was done by the FBI Laboratory in Quantico, VA (Brandhagen et al., 2020). This publication describes the developmental and internal validation studies done for mtDNA casework. It is important to mention that, as explained in the FBI's QAS, internal validation studies can supplement any validation study that is deficient, or, if conducted within the same laboratory, developmental validation work may satisfy some of the internal validation studies. The FBI's work focused on testing and implementing the PowerSeq CRM Nested System (Promega, WI), an NGS assay that generates information on the mtDNA control region, into the FBI Laboratory's operational casework. This NGS mtDNA system was combined with the Illumina MiSeq instrument for sequencing and the GeneMarker HTS software package (SoftGenetics, State College, PA) for data analysis. All studies were conducted in accordance with the SWGDAM validation guidelines for forensic DNA analysis methods and the FBI's QAS for forensic DNA testing laboratories and were submitted for NDIS approval. As already mentioned in this chapter, as of May 2019, the control region data generated using the PowerSeq CRM Nested System is approved for upload to the DNA database. The results obtained in this study demonstrated that this mtDNA NGS system is efficient, accurate, reproducible, sensitive, and reliable. According to the authors, a major advantage of this assay encountered during the validation process was its high sensitivity. Complete control region sequence data was obtained from as few as 2000 copies of mtDNA. To put this into perspective, there are 2–10 copies of mtDNA in one mitochondrion, and there are up to 1000 mitochondria per cell. Taking into consideration the mtDNA quantitation values routinely encountered in the FBI Laboratory's casework,

the authors anticipate an increase in the number of complete control region haplotypes generated from questioned samples. It is expected that control region data will be retrieved from approximately 90% of routine FBI casework samples, which is much more than the 20% with the current Sanger sequencing-based protocols. The studies also demonstrated that only 30% of the extract volume is required to generate profiles from degraded samples when compared to Sanger sequencing. The authors hope that these published studies can provide a model for any forensic laboratory seeking to validate NGS protocols for mtDNA typing according to current SWGDAM guidelines.

Funding for forensic laboratories to adopt and validate NGS technology

As already mentioned at the beginning of this chapter, the adoption of NGS by forensic laboratories has been slow, and the high cost of instrumentation and the validation and training processes are often barriers to the implementation of this new technology in forensic analysis. A question that is frequently asked is if there are funds available for the implementation of NGS. In the United States, there are generally two ways for a forensic laboratory to get funding. The first option is through allocated budgets by federal, state, or local annual budgets, and the second is through grants.

Federal funding is usually offered by the US Department of Justice, and one of the opportunities that accredited forensic laboratories can apply for is the DNA Capacity Enhancement for Backlog Reduction (CEBR) Program. For the fiscal year of 2021, more than 80 million dollars were available to fund crime labs and help them reduce and/or prevent a backlog of forensic and database DNA samples (Bureau of Justice Assistance, U.S. Department of Justice, 2021). Through this funding program, laboratories can purchase newer and more efficient instruments as well as evaluate, validate, and implement new chemistries that are approved for use with NDIS, which now include NGS systems. Another funding opportunity is the Paul Coverdell National Forensic Science Improvement Act (NFSIA) Program (Bureau of Justice Assistance, US Department of Justice, n.d.a), which has part of the funding program allocated to address emerging forensic science technology, such as new types of instrumentation. These two programs are known as formula grants, and they are awarded based on state crime statistics and/or population.

Other funding opportunities are also available, and these are known as competitive grants, which are evaluated by an external review panel and are not given just to accredited crime laboratories.

A popular grant opportunity is the National Institute of Justice (NIJ) Forensic Science Research and Development Program (National Institute of Justice, n.d.). This program finds basic or applied research and development projects that include:

- physical capital by supporting the acquisition, maintenance, and development of laboratory instrumentation.

- intellectual capital by supporting researchers and providing learning and training experiences for scientists at all career stages.
- structural capital by funding projects that support databases and add to the scientific literature.

Another competitive grant is the NIJ Postconviction Testing of DNA Evidence Grant Program (Bureau of Justice Assistance, US Department of Justice, n.d.b), which provides funding to states for postconviction DNA testing for violent felony offenses (as defined by state law) in which actual innocence might be demonstrated. This grant cannot be used for laboratory equipment but can be allocated to the purchase of laboratory supplies, reagents, software programs, and personnel to locate and analyze biological evidence.

When searching for external funding or in the scenario where funding for validation is available in the annual budget, forensic laboratories should take into consideration their needs, infrastructure, and personnel availability when deciding which NGS system to adopt and consequentially validating and establishing SOPs and workflows.

References

Alonso, A., Müller, P., Roewer, L., Willuweit, S., Budowle, B., & Parson, W. (2017). European survey on forensic applications of massively parallel sequencing. *Forensic Science International Genetics, 29*, e23–e25. https://doi.org/10.1016/j.fsigen.2017.04.017

Andrews, R. M., Kubacka, I., Chinnery, P. F., Lightowlers, R. N., Turnbull, D. M., & Howell, N. (1999). Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nature Genetics, 23*(2), 147.

Ballard, D., Winkler-Galicki, J., & Wesoły, J. (2020). Massive parallel sequencing in forensics: Advantages, issues, technicalities, and prospects. *International Journal of Legal Medicine, 134*(4), 1291–1303. https://doi.org/10.1007/s00414-020-02294-0

Brandhagen, M. D., Just, R. S., & Irwin, J. A. (2020). Validation of NGS for mitochondrial DNA casework at the FBI Laboratory. *Forensic Science International. Genetics, 44*, 102151. https://doi.org/10.1016/j.fsigen.2019.102151

Bruijns, B., Tiggelaar, R., & Gardeniers, H. (2018). Massively parallel sequencing techniques for forensics: A review. *Electrophoresis, 39*(21), 2642–2654. https://doi.org/10.1002/elps.201800082

Buchard, A., Kampmann, M.-L., Poulsen, L., Børsting, C., & Morling, N. (2016). ISO 17025 validation of a next-generation sequencing assay for relationship testing. *Electrophoresis, 37*(21), 2822–2831. https://doi.org/10.1002/elps.201600269

Bureau of Justice Assistance, U.S. Department of Justice. (n.d.a). Paul Coverdell National Forensic Science Improvement Act (NFSIA) Program. Retrieved March 2022, from https://bja.ojp.gov/program/coverdell/funding.

Bureau of Justice Assistance, U.S. Department of Justice. (n.d.b). Postconviction testing of DNA evidence. Retrieved March 2022, from https://bja.ojp.gov/program/postconviction-testing-dna-evidence/overview.

Bureau of Justice Assistance, U.S. Department of Justice. (2021). *BJA FY 21 DNA capacity enhancement for backlog reduction (CEBR) program*. https://bja.ojp.gov/sites/g/files/xyckuh186/files/media/document/O-BJA-2021-128001.pdf.

Butler, J. (2012). Advanced topics. In *Forensic Dna typing: Methodology*. Elsevier Academic Press. https://doi.org/10.1016/C2011-0-04189-3

Casals, F., Anglada, R., Bonet, N., Rasal, R., van der Gaag, K. J., Hoogenboom, J., Solé-Morata, N., Comas, D., & Calafell, F. (2017). Length and repeat-sequence variation in 58 STRs and 94 SNPs in

two Spanish populations. *Forensic Science International. Genetics, 30*, 66–70. https://doi.org/10.1016/j.fsigen.2017.06.006

Cellmark Forensic Services. (n.d.). DNA services. Retrieved February 2022, from https://www.cellmarkforensics.co.uk/services/dna-services/.

Cihlar, J. C., Amory, C., Lagacé, R., Roth, C., Parson, W., & Budowle, B. (2020). Developmental validation of a MPS workflow with a PCR-based short amplicon whole mitochondrial genome panel. *Genes, 11*(11). https://doi.org/10.3390/genes11111345

European Network of Forensic Science Institutes (ENFSI). (2010). *Recommended minimum criteria for the validation of various Aspects of the DNA profiling process*. http://enfsi.eu/wp-content/uploads/2016/09/minimum_validation_guidelines_in_dna_profiling_-_v2010_0.pdf.

Faccinetto, C., Sabbatini, D., Serventi, P., Rigato, M., Salvoro, C., Casamassima, G., Margiotta, G., De Fanti, S., Sarno, S., Staiti, N., Luiselli, D., Marino, A., & Vazza, G. (2021). Internal validation and improvement of mitochondrial genome sequencing using the precision ID mtDNA whole genome panel. *International Journal of Legal Medicine, 135*(6), 2295–2306. https://doi.org/10.1007/s00414-021-02686-w

Federal Bureau of Investigation (FBI). (2020). *Guidance document for quality assurance standards for forensic DNA testing and DNA databasing laboratories*. https://www.fbi.gov/file-repository/qas-guidance-document-070120.pdf/view.

Federal Bureau of Investigation (FBI). (2021). *National DNA index system (NDIS) operational procedures manual*. https://www.fbi.gov/file-repository/ndis-operational-procedures-manual.pdf/view.

Federal Bureau of Investigation (FBI). (n.d.). CODIS and NDIS fact sheet. Retrieved January 2022, from https://www.fbi.gov/services/laboratory/biometric-analysis/codis/codis-and-ndis-fact-sheet.

Finaughty, C., Reid, K. M., Alli, I. H., & Heathfield, L. J. (2020). A first for forensic genetics in Africa: Successful identification of skeletal remains from the marine environment using massively parallel sequencing. *Forensic Science International. Genetics, 49*, 102370. https://doi.org/10.1016/j.fsigen.2020.102370

Gettings, K. B., Ballard, D., Bodner, M., Borsuk, L. A., King, J. L., Parson, W., & Phillips, C. (2019). Report from the STRAND Working Group on the 2019 STR sequence nomenclature meeting. *Forensic Science International. Genetics, 43*, 102165. https://doi.org/10.1016/j.fsigen.2019.102165

Gettings, K. B., Borsuk, L. A., Steffen, C. R., Kiesler, K. M., & Vallone, P. M. (2018). Sequence-based U.S. population data for 27 autosomal STR loci. *Forensic Science International. Genetics, 37*, 106–115. https://doi.org/10.1016/j.fsigen.2018.07.013

Hollard, C., Ausset, L., Chantrel, Y., Jullien, S., Clot, M., Faivre, M., Suzanne, É., Pène, L., & Laurent, F.-X. (2019). DNA Signature Preparation kit for high-throughput analysis in forensic laboratories. *Forensic Science International. Genetics, 40*, 37–45. https://doi.org/10.1016/j.fsigen.2019.01.010

Hussing, C., Bytyci, R., Huber, C., Morling, N., & Børsting, C. (2019). The Danish STR sequence database: Duplicate typing of 363 Danes with the ForenSeq™ DNA Signature Prep kit. *International Journal of Legal Medicine, 133*(2), 325–334. https://doi.org/10.1007/s00414-018-1854-0

Jäger, A. C., Alvarez, M. L., Davis, C. P., Guzmán, E., Han, Y., Way, L., Walichiewicz, P., Silva, D., Pham, N., Caves, G., Bruand, J., Schlesinger, F., Pond, S. J. K., Varlaro, J., Stephens, K. M., & Holt, C. L. (2017). Developmental validation of the MiSeq FGx forensic Genomics system for targeted next generation sequencing in forensic DNA casework and database laboratories. *Forensic Science International. Genetics, 28*, 52–70. https://doi.org/10.1016/j.fsigen.2017.01.011

Köcher, S., Müller, P., Berger, B., Bodner, M., Parson, W., Roewer, L., & Willuweit, S. (2018). Interlaboratory validation study of the ForenSeq™ DNA Signature Prep Kit. *Forensic Science International. Genetics, 36*, 77–85. https://doi.org/10.1016/j.fsigen.2018.05.007

Laurent, F.-X. (2018). *Implementation of a fully automated workflow using MPS technology in a high volume casework laboratory*. https://cdn.asp.events/CLIENT_CL_ME_3820BA96_5056_B731_4C48B7366FE2A35A/sites/Forensics-Africa-2020/media/GCC-Forensic-Conference-2018–-FX-Laurent.pdf.

Moura-Neto, R., King, J. L., Mello, I., Dias, V., Crysup, B., Woerner, A. E., Budowle, B., & Silva, R. (2021). Evaluation of Promega PowerSeq™ Auto/Y systems prototype on an admixed sample of Rio de Janeiro, Brazil: Population data, sensitivity, stutter and mixture studies. *Forensic Science International. Genetics, 53*, 102516. https://doi.org/10.1016/j.fsigen.2021.102516

National Institute of Justice. (n.d.). Forensic science research and development. Retrieved March 2022, from https://nij.ojp.gov/topics/forensics/forensic-science-research-and-development.

National Institute of Standards and Technology (NIST). (n.d.). Validation information to aid forensic DNA laboratories. Retrieved January 2022, from https://strbase.nist.gov/validation.htm.

Novroski, N. M. M., King, J. L., Churchill, J. D., Seah, L. H., & Budowle, B. (2016). Characterization of genetic sequence variation of 58 STR loci in four major population groups. *Forensic Science International. Genetics, 25*, 214–226. https://doi.org/10.1016/j.fsigen.2016.09.007

Parson, W., Ballard, D., Budowle, B., Butler, J. M., Gettings, K. B., Gill, P., Gusmão, L., Hares, D. R., Irwin, J. A., King, J. L., Knijff, P., de, Morling, N., Prinz, M., Schneider, P. M., Neste, C. V., Willuweit, S., & Phillips, C. (2016). Massively parallel sequencing of forensic STRs: Considerations of the DNA commission of the International Society for Forensic Genetics (ISFG) on minimal nomenclature requirements. *Forensic Science International. Genetics, 22*, 54–63. https://doi.org/10.1016/j.fsigen.2016.01.009

Phillips, C., Devesse, L., Ballard, D., van Weert, L., de la Puente, M., Melis, S., Álvarez Iglesias, V., Freire-Aradas, A., Oldroyd, N., Holt, C., Syndercombe Court, D., Carracedo, Á., & Lareu, M. V. (2018). Global patterns of STR sequence variation: Sequencing the CEPH human genome diversity panel for 58 forensic STRs using the Illumina ForenSeq DNA Signature Prep Kit. *Electrophoresis, 39*(21), 2708–2724. https://doi.org/10.1002/elps.201800117

Phillips, C., Gettings, K. B., King, J. L., Ballard, D., Bodner, M., Borsuk, L., & Parson, W. (2018b). "The devil's in the detail": Release of an expanded, enhanced and dynamically revised forensic STR Sequence Guide. *Forensic Science International. Genetics, 34*, 162–169. https://doi.org/10.1016/j.fsigen.2018.02.017

Promega. (n.d.a). PowerSeq® 46GY system. Retrieved January 2022, from https://www.promega.com/products/forensic-dna-analysis-mps/target-amplification-and-library-prep/powerseq-46gy-system/?catNum=PS4600.

Promega. (n.d.b). PowerSeq CRM nested system, custom. Retrieved January 2022, from https://www.promega.com/products/forensic-dna-analysis-mps/target-amplification-and-library-prep/powerseq-crm-nested-system-custom/?catNum=AX5810.

Scientific Working Group for DNA Analysis Methods (SWGDAM). (2016). *Validations guidelines for DNA analysis methods*. https://docs.wixstatic.com/ugd/4344b0_813b241e8944497e99b9c45b163b76bd.pdf.

Scientific Working Group for DNA Analysis Methods (SWGDAM). (2019a). *Addendum to "SWGDAM interpretation guidelines for autosomal STR typing by forensic DNA testing laboratories" to address next generation sequencing*. https://www.swgdam.org/_files/ugd/4344b0_91f2b89538844575a9f51867def7be85.pdf.

Scientific Working Group for DNA Analysis Methods (SWGDAM). (2019b). *Interpretation guidelines for mitochondrial DNA analysis by forensic DNA testing laboratories*. https://www.swgdam.org/_files/ugd/4344b0_f61de6abf3b94c52b28139bff600ae98.pdf.

Scientific Working Group for DNA Analysis Methods (SWGDAM). (2021). *Interpretation guidelines for autosomal STR typing by forensic DNA testing laboratories*. https://www.swgdam.org/_files/ugd/4344b0_3f94c9a6286048c3924c58e2c230e74e.pdf.

Scientific Working Group for DNA Analysis Methods (SWGDAM). (n.d.). Retrieved January 2022, from https://www.swgdam.org.

Senst, A., Caliebe, A., Scheurer, E., & Schulz, I. (2022). Validation and beyond: Next generation sequencing of forensic casework samples including challenging tissue samples from altered human corpses using the MiSeq FGx system. *Journal of Forensic Sciences*. https://doi.org/10.1111/1556-4029.15028

Silva, D. S. B. S., Sawitzki, F. R., Scheible, M. K. R., Bailey, S. F., Alho, C. S., & Faith, S. A. (2018). Genetic analysis of Southern Brazil subjects using the PowerSeq™ AUTO/Y system for short tandem repeat sequencing. *Forensic Science International. Genetics, 33*, 129–135. https://doi.org/10.1016/j.fsigen.2017.12.008

Silva, D. S. B. S., Scheible, M. K., Bailey, S. F., Williams, C. L., Allwood, J. S., Just, R. S., Schuetter, J., Skomrock, N., Minard-Smith, A., Barker-Scoggins, N., Eichman, C., Meiklejohn, K., & Faith, S. A. (2020). Sequence-based autosomal STR characterization in four US populations using PowerSeq™ Auto/Y system. *Forensic Science International. Genetics, 48*, 102311. https://doi.org/10.1016/j.fsigen.2020.102311

Silvery, J., Ganschow, S., Wiegand, P., & Tiemann, C. (2020). Developmental validation of the monSTR identity panel, a forensic STR multiplex assay for massively parallel sequencing. *Forensic Science International. Genetics, 46*, 102236. https://doi.org/10.1016/j.fsigen.2020.102236

Thermo Fisher Scientific. (n.d.). Precision ID NGS system for human identification: Targeted sequencing for forensic DNA analysis. Retrieved February 2022, from https://www.thermofisher.com/us/en/home/industrial/forensics/human-identification/forensic-dna-analysis/dna-analysis/next-generation-sequencing-ngs-forensics.html.

Thermo Fisher Scientific Applied Biosystems. (n.d.). Expand your forensic workflow with the precision ID NGS system. Retrieved January 2022, from https://assets.thermofisher.com/TFS-Assets/GSD/brochures/precisionid-ngs-system-brochure.pdf.

van der Gaag, K. J., de Leeuw, R. H., Hoogenboom, J., Patel, J., Storts, D. R., Laros, J. F. J., & de Knijff, P. (2016). Massively parallel sequencing of short tandem repeats-population data and mixture analysis results for the PowerSeq^{TM} system. *Forensic Science International. Genetics, 24*, 86–96. https://doi.org/10.1016/j.fsigen.2016.05.016

Verogen. (2019a). *Cellmark and Verogen collaborate to make the next generation of forensic DNA analysis*. Available to UK Police, https://verogen.com/cellmark-and-verogen-bring-next-gen-forensic-dna-to-uk/.

Verogen. (2019b). *FBI approves Verogen's next-gen forensic DNA technology for national DNA index system (NDIS)*. https://verogen.com/ndis-approval-of-miseq-fgx/.

Verogen. (n.d.a). ForenSeq DNA Signature Prep Kit. Retrieved January 2022, from https://verogen.com/products/forenseq-dna-signature-prep-kit/.

Verogen. (n.d.b). ForenSeq Kintelligence Kit. Retrieved January 2022, from https://verogen.com/products/forenseq-kintelligence-kit/.

Verogen. (n.d.c). ForenSeq MainstAY Product Line. Retrieved January 2022, from https://verogen.com/products/forenseq-mainstay-kit/.

Verogen. (n.d.d). ForenSeq mtDNA Control Region Kit. Retrieved January 2022, from https://verogen.com/products/forenseq-mtdna-control-region-kit/.

Verogen. (n.d.e). ForenSeq mtDNA Whole Genome Kit. Retrieved January 2022, from https://verogen.com/products/forenseq-mtdna-whole-genome-kit/.

Verogen. (n.d.f). How next generation sequencing resolved a difficult case, leading to the first criminal conviction of its kind. Retrieved February 2022, from https://verogen.com/wp-content/uploads/2020/12/ngs-first-criminal-conviction-case-study-vd2019024-b.pdf.

Wu, J., Li, J.-L., Wang, M.-L., Li, J.-P., Zhao, Z.-C., Wang, Q., Yang, S.-D., Xiong, X., Yang, J.-L., & Deng, Y.-J. (2019). Evaluation of the MiSeq FGx system for use in forensic casework. *International Journal of Legal Medicine, 133*(3), 689–697. https://doi.org/10.1007/s00414-018-01987-x

This page intentionally left blank

CHAPTER 26

Challenges in using genetic genealogy in forensics

Praveen Kumar Yadav[1] **and Sweety Sharma**[2]
[1]Centre for Integrated Studies, Cochin University of Science and Technology, Kochi, Kerala, India; [2]Department of Forensic Biology and Biotechnology, School of Forensic Sciences, LNJN National Institute of Criminology and Forensic Science, National Forensic Science University, New Delhi, India

Introduction

Forensic genetic genealogy, or investigative genetic genealogy (IGG), emerged as a novel technique in 2018 that combines forensic genetics with conventional genealogy and genetics (Glynn, 2022). Genetic genealogy has traditionally been used to identify family history or to regain the genetic heritage of a person who unfortunately lost their biological identity through abandonment, adoption, misattribution of parentage, gamete donation, etc. Recently, this technique has found its way into the criminal investigation system, where it is being used to find relatives of perpetrators through shared DNA. In 2018, IGG was initially used for the identification of Joseph James DeAngelo, who was the prime suspect in the famous Golden State killer case. This case was highly publicized in print as well as online media. Since then, this technique has garnered global attention and is continuously being used in the investigation process by forensic and law enforcement professionals. Fig. 26.1 illustrates the concept of genealogy in familial searching and shows how a suspect can be connected to close (child-parents; descendants) as well as distant (first cousin and beyond) relative members through shared DNA (Berkman et al., 2018; Deb, 2022; Greytak et al., 2019; Thomson et al., 2020).

Fig. 26.2 shows the basic overview of IGG from the genealogist's point of view. In a nutshell, DNA-containing samples are recovered from the crime scene. Using SNP profiling, questionable genomic data is generated for this sample. This questionable genomic data is then run through the genealogical databases. The software compares this questioned genomic data with the profiles present in the database, and suitable matches are then suggested. These possible matches are then used as investigative leads to find out the perpetrators.

How forensic DNA profiling differs from forensic genetic genealogy

In principle, both of these techniques are similar; however, in methodology and application, they vary significantly. These differences depend on various factors such as the DNA

Next Generation Sequencing (NGS) Technology in DNA Analysis
ISBN 978-0-323-99144-5, https://doi.org/10.1016/B978-0-323-99144-5.00007-X
© 2024 Elsevier Inc.
All rights reserved.

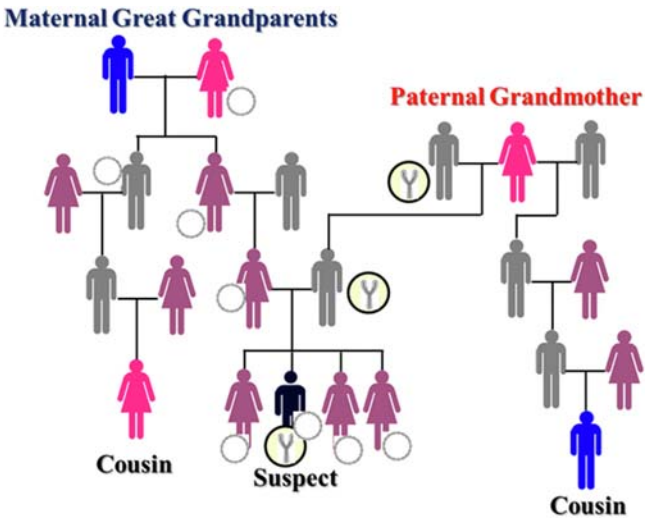


Figure 26.1 Pedigree showing relatedness. *(Image source: sciencenews.org.)*

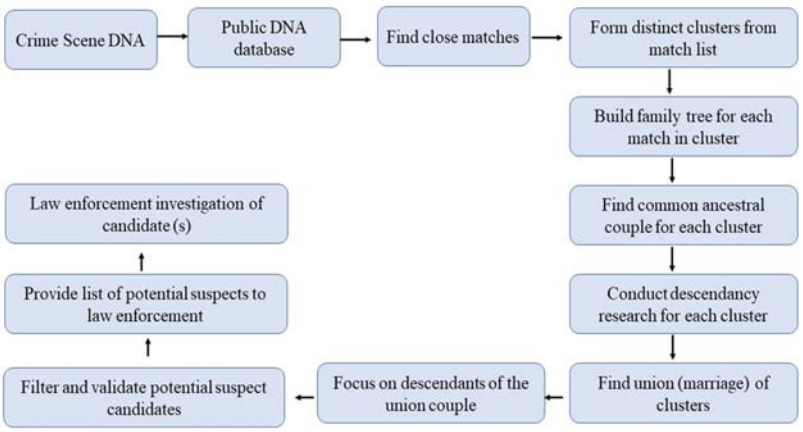

Crime Scene DNA
Public DNA database
Find close matches
Form distinct clusters from match list
Build family tree for each match in cluster
Find common ancestral couple for each cluster
Conduct descendancy research for each cluster
Find union (marriage) of clusters
Focus on descendants of the union couple
Filter and validate potential suspect candidates
Provide list of potential suspects to law enforcement
Law enforcement investigation of candidate (s)

Figure 26.2 Basic overview of IGG from the genealogist's point of view (McDermott, 2020).

database searched, the data generated, the genomic region used, the technology used, and the types and number of DNA markers used (Table 26.1).

How investigative genetic genealogy works

To generate the investigative leads, genetic similarities among the known familial relations are used. The basic information used in the IGG process is divided into two categories: (i) Genetic relative information, which can be generated using a genetic genealogy

Table 26.1 Summary of differences for FDP and FGG.

	Databases	DNA markers	Region of genome	No.of markers	Latest technology	Data file generated
FDP	National DNA databases (US-CODIS core 20 STRs; UK and Ireland employing DNA-17)	STRs	Noncoding	16−27	CE and PCR	Electropherogram
FGG	Genetic genealogical databases (FamilyTreeDNA, GEDmatch PRO, DNA solves etc.)	SNPs	Coding	∼600,000 and ∼1 million	Next-generation sequencing and whole genome Sequencing Targeted SNP kits	FASTQ

From Glynn, C. (2022). *Genes*, 13, 1381. https://doi.org/10.3390/genes13081381.

database based on its internal comparison of SNP profiles in the database; and (ii) Genealogical and other frequently publicly available information from eulogies, census records, and public databases. The genealogical information is then used to depict relationships and generate a list of relatives. Genetic relatedness is deduced using DNA segments that are shared by identical by descent (IBD) individuals; that means every individual has inherited segments matching with common ancestors. Further, this information is utilized by enforcement agencies to develop the family tree and identify high-likelihood suspects. The complete process of IGG (steps 1 and 2) that is book-ended by standard police work is illustrated in Fig. 26.3 (Guerrini et al., 2021; McDermott, 2020).

IGG is an emerging technology that is most commonly used in the United States (US) and European countries. The emergence of many direct-to-consumer (DTC) databases has made the required genomic information easily accessible to investigative agencies,

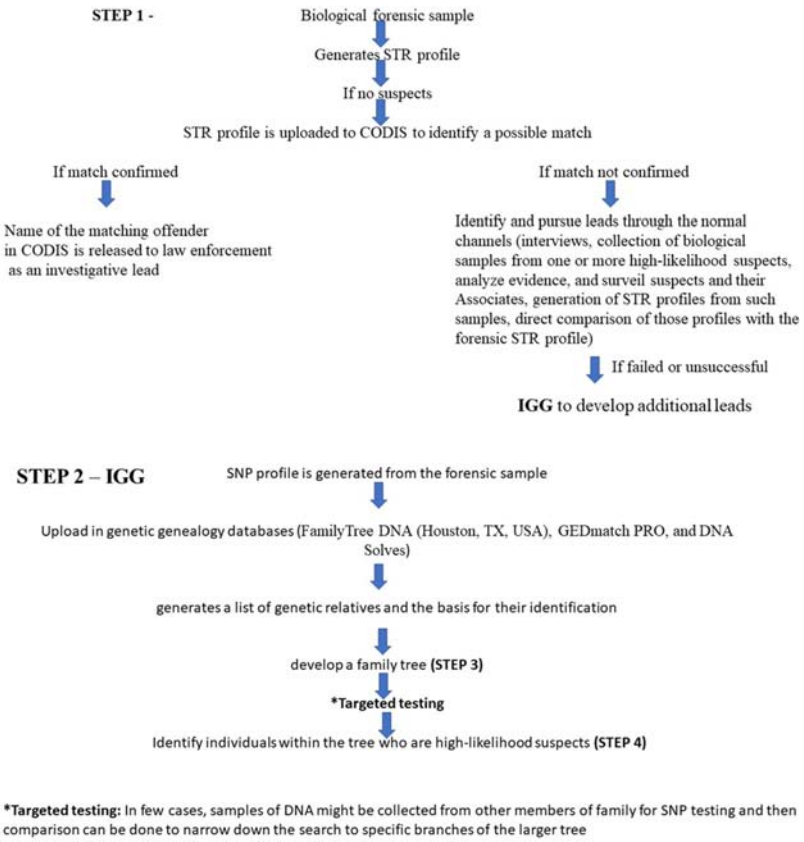

STEP 1 -
Biological forensic sample
Generates STR profile
If no suspects
STR profile is uploaded to CODIS to identify a possible match
If match confirmed
Name of the matching offender in CODIS is released to law enforcement as an investigative lead
If match not confirmed
Identify and pursue leads through the normal channels (interviews, collection of biological samples from one or more high-likelihood suspects, analyze evidence, and surveil suspects and their Associates, generation of STR profiles from such samples, direct comparison of those profiles with the forensic STR profile)
If failed or unsuccessful
IGG to develop additional leads
STEP 2 – IGG
SNP profile is generated from the forensic sample
Upload in genetic genealogy databases (FamilyTree DNA (Houston, TX, USA), GEDmatch PRO, and DNA Solves)
generates a list of genetic relatives and the basis for their identification
develop a family tree (STEP 3)
*Targeted testing
Identify individuals within the tree who are high-likelihood suspects (STEP 4)
*Targeted testing: In few cases, samples of DNA might be collected from other members of family for SNP testing and then comparison can be done to narrow down the search to specific branches of the larger tree

Figure 26.3 Basic steps involved in investigative genetic genealogy (IGG).

which has also added to its overwhelming use in the investigation process. Countries such as the UK, Sweden, Australia, England, the Netherlands, Ireland, and Germany are also exploring its potential use to identify unidentified human remains and solve violent crimes (Mehar, 2021; Samuel & Kennett, 2020; Scudder et al., 2020; Thomson et al., 2020; Tillmar et al., 2021).

Genetic genealogical databases such as Oxford Ancestors (Bicester, UK; year 2000), FamilyTreeDNA (Houston, TX, USA; year 2000; www.familytreedna.com), GEDmatch PRO (2010; www.gedmatch.com), DNA Solves, 23andMe (South San Francisco, CA, USA; year 2007; www.23andme.com), MyHeritage (Or Yehuda, Israel; year 2016; www.myheritage.com/dna), AncestryDNA (Lee Hay, UT, USA; year 2012), and Ancestry.com (1996; www.ancestry.com/dna) gather genomic data, while some of these allow the information to be used for investigative purposes. As of July 2022, the top genealogical databases have 41 million registered users, out of which ~21 million are in AncestryDNA, ~12.8 million in 23andMe, ~6 million in MyHeritageDNA, and ~1.77 million in FamilyTreeDNA (Glynn, 2022; Mehar, 2021). This number continues to grow rapidly as IGG gains more acclaim after solving a series of famous cases (Greytak et al., 2019). This has also sparked an intense debate on the challenges and ethical considerations related to the extensive use of IGG. Although IGG has provided an efficient tool for solving new and cold cases, it has also raised many ethical and social concerns that need to be addressed. It is imperative that a solution be found to these challenges to make the use of IGG more transparent, ethical, and socially acceptable.

Challenges in using genetic genealogy

Errors

SNP genotypes can experience two types of errors: (a) Technological errors and (b) induced errors (Fig. 26.4). Technological errors are those arising during the process of amplification and sequencing or during the bioinformatics pipeline, which includes sequence alignment or variant calling (Kling et al., 2021). Once the sequencing has been completed a large amount of data is generated, which requires it to be aligned with the reference data before comparison. This is challenging as the read lengths are tiny, ranging from 36 to 250 bps. The presence of multiple similar sites, repetitive zones, pseudogenes, and a massive volume of the data generated can make this process tedious and challenging (Duarte Castelão, 2017). Studies report that genotyping errors can be less than 1%, whereas phasing errors can be around 0.64% (Browning & Browning, 2007; Durand et al., 2014). These errors can significantly affect the likelihood ratio calculations up to 0.05% (Kling, 2019). One method to reduce these types of errors is to use a haplotype score incorporating the phase and genotype error rates, which helps in filtering the spurious IBD segments during the postprocessing steps (Durand et al., 2014). Some other methods to reduce these errors include incorporating the error rates into the segment

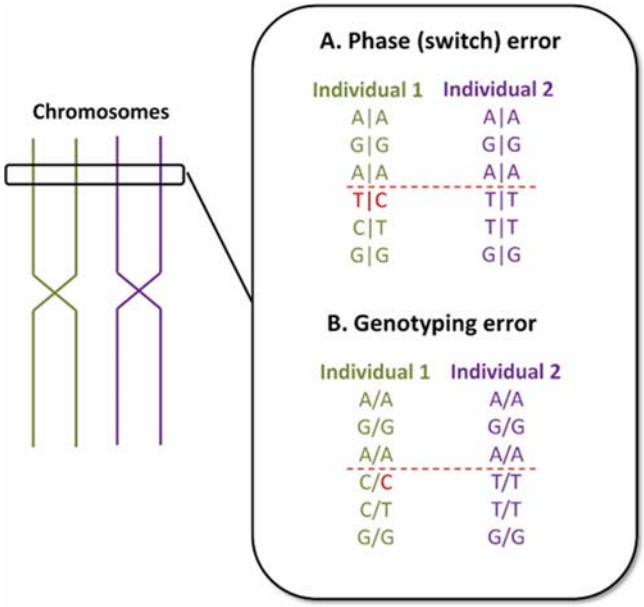


Figure 26.4 Illustration of two types of errors. An IBD segment is prematurely terminated at the dashed red line with errors highlighted in red. (A) Phase errors occur when the chromosomes of an individual are separated into maternal and paternal origins and where the process of phasing (wrongly) switches chromosomes. (B) Genotyping errors occur either during the amplification process or at the bioinformatic genotyping level. *(From Edge, M., & Coop, G. (2020).* eLife*, 9. https://doi.org/10.7554/eLife.51810.)*

models (Browning & Browning, 2013; Dimitromanolakis et al., 2019) or modeling for the errors themselves (Ball et al., 2016; Henn et al., 2012).

Samples available for database sequencing are of better quality and quantity; therefore, a complete genome is available for exploration (high-depth genome sequencing). The forensic samples are often of inferior quality and quantity when low-depth genome sequencing is deployed. Due to this, higher error rates are observed when the questioned sample is searched in the database (Saunders et al., 2007).

DNA mixtures

Databases are built using single-source DNA; however, DNA from multiple sources is common in forensic cases. There are well-established protocols for the examination of such samples and for calculating their evidentiary values (Bleka et al., 2016; Bright et al., 2016; Haned, 2011). However, extending the examination to genealogy might be challenging. Searches must be conducted using a single-source DNA profile (Phillips, 2018). This can be accomplished by separating the profiles of contributors either through conditioning or by developing a statistical model for it. However, in doing so, a level of

uncertainty is generated, which can then affect the false/true positive rates (Bradford et al., 2011). Low-quality DNA from forensic cases yields very less information, even when whole genome sequencing is employed; however, a statistical model could extract contributors from the mixture based on the allele dosage, also known as read counts. In the case of mixtures with homozygote genotypes, one can say that the perpetrator will also be homozygote. However, in the case of a heterozygote, the genotype can be homozygous or heterozygous for the allele. This leads to a rise in the number of false positives (Kling et al., 2021).

Degraded or low-quality forensic samples

In next-generation sequencing, the quality of extracted DNA depends on yield and degradation index (DI). Good-quality DNA has an A_{260}/A_{280} ratio of 1.7−1.8. Lower-ratio DNA might be suitable for other applications, but only after some pretreatment to remove the contaminants. A higher yield indicates good quality, but this is not always the case. Sometimes, despite a high yield, the quality of the extracted DNA might be poor. This is when the extracted DNA sample contains a large quantity of junk DNA (Davawala et al., 2022). Another factor to consider here is the DI. The DI is the ratio of smaller to larger DNA fragments in a given sample. A low-yield DNA sample with a low DI is better than a high-yield DNA sample with a high DI.

Need for standards

There is a consensus among the IGG practitioners that there is a lack of official standards and certifications. Gurney et al. reported the formation of a nonprofit Board of Certification for IGG, comprising the most renowned IGG practitioners in the USA, to develop IGG standards (Gurney et al., 2022). Some other available guidelines include the Interim Policy on Forensic Genetic Genealogical DNA Analysis and Searching (released in November 2019 by the US Department of Justice) (Callaghan, 2019) and the overview of IGG (released in February 2020 by the Scientific Working Group on DNA Analysis Methods, SWGDAM) (Epstein et al., 2000). Still, there is an immense need to develop standardized methods, policies, and protocols for the application of genetic genealogy in the field of criminal investigations.

IGG has developed as a specialization of genetic genealogy where the methodologies are similar. However, their impact on people might be significantly different. An individual who seeks to find out his/her relatives might give their consent; however, the same person might retract his/her consent if the same information was to be used in a criminal investigation (Gurney et al., 2022). Standardizing the techniques, methodologies, and their applications raises public trust against their misuse. This further helps in reducing the public's fear while consenting to the use of their genealogy in a criminal investigation. The public also fears that IGG practitioners might be ignoring the terms of service. Standards also subside these public concerns. Ram et al. (2021) words that "law enforcement

privacy advocates found value in developing guardrails [for IGG]", the latter because "limits on the technology (of IGG) to promote public trust would help ensure that the technology remains available to assist them in their work," seem to summarize this efficiently.

As the public must have trust in IGG practitioners, the investigating agencies must also have trust in the IGG practitioner's qualifications and ability to perform the assiduous tasks of IGG. Standardization could achieve this by controlling the proficiency and ethical level delivered by the IGG practitioners. Proficient practitioners can provide value for the money, time, and other resources spent on their services, which otherwise go to waste. Also, subpar and unethical performances can jeopardize the investigation, in which case agencies might totally abandon its use despite its huge potential to provide important leads in criminal investigations. To maintain professional proficiency, accreditation standards can be used. Standards can also help in setting up accountability where proper protocols and standards are not followed. For these reasons, standards can help build the trust of the public and the investigating agencies.

Direct-to-customer genomic services

In 2021, the DTC genomic industry was valued at around USD 1.29 billion (Direct-To-Consumer Genetic Testing Market, 2022), which increased to more than double the market value of USD 3 billion in 2022. It stands to grow at an 11.5% compound annual growth rate from 2023 to 2032 and reach a market value higher than USD 10.5 billion (Swain & Subodh, 2023). This ballooning growth is due to the decreasing cost of large-scale SNP typing, rising public awareness about DTC services, rising genetic disorders, increased interest in finding relatives and ancestors, and growing demand for such services (Court, 2018; Naveed et al., 2015). Most of these services can be availed of through online platforms (around 77%), whereas the remaining 23% of DTC services are over-the-counter services. North American and European countries continue to be the major consumers of DTC genomic services. In the forensic context, there are several issues pertaining to DTC genomic services. These include false identification, the use of unproven or invalid testing methods, and biospecimen or hacking threats (Apathy et al., 2018; Carroll et al., 2020; Salloum et al., 2018).

False identification

Using DTC genomic services to find relatives can be challenging and misleading, especially as you move far back or far across the pedigree. The severity of the problem is aggravated because a person could have 1000 fourth cousins and around 5000 fifth cousins, out of which only a few could be included in the database. The IGG practitioners use a triangulation method for identifying the parts of shared genetic sequences between the perpetrator and the database; however, using DNA as the sole source might be less fruitful at best and erroneous at worst. Therefore, it is advisable to augment the

results obtained from the DTC genomic search using conventional genealogical methods (Court, 2018).

Use of unproven or invalid testing methods

There is a severe lack of genealogical standards in the market, and it is highly pertinent that the use of genealogical services be regulated. In the absence of such standards, it is very difficult to validate the results provided by these DTC genomic services. Many studies have raised serious concerns about the validity of the methods used by these DTC services (Court, 2018; Naveed et al., 2015).

Biospecimen and hacking threats

DNA is extracted from the cells in its chemical form, and then, using computing systems it is digitized (Naveed et al., 2015). Sensitive information pertaining to the genotypic and phenotypic makeup is stored in the databases. Owing to the digital nature of databases, they are susceptible to cyberattacks and information loss. Similar attacks have been reported on fingerprint repositories (Aldhous, 2020). In July 2020, genomic data of 1 million users was leaked from two online databases, GEDMatch and MyHeritage using a phishing email, redirecting to a fake login page, and then collecting the username and password. Any attack on a similar database can severely diminish its credibility and stigmatize the usage of genomic databases among people as well as investigative agencies.

Intrusion to privacy

One of the most important privacy-related questions raised by Wickenheiser is: To what extent can the rights of the innocent public and relatives of the perpetrator of a crime be infringed upon by interrogating their genetic data to identify the crime perpetrator, prevent future crimes, and improve public safety? Using genealogical surveillance can help in apprehending criminals even where no hit is found in the criminal DNA databases. However, it raises a serious concern about the individual's freedom, right to privacy, and autonomy even when their DNA is not found at the crime scene. Practitioners, law enforcement agencies, and police personnel access personal information, including age, sex, and location, about both living and dead individuals (Gurney et al., 2022).

Genomic privacy

Genomic data is not exceptional; however, when interpreted properly, it can provide sensitive information about any individual. This data-centric information has widespread applications, and its leakage can have immense implications in terms of discrimination in insurance, employment, education, etc. One such example was shared by Dr. Noralane Lindor at the Mayo Clinic's Individualizing Medicine Conference (Lindor et al., 2013). She states that while studying cancer patients, she also sequenced the grandchildren of

one of her patients and found that two of the grandchildren had the same mutation as the patient. When one grandchild applied for the United States Army for the position of helicopter pilot, she was rejected even though the position does not require any genetic testing. Many other examples of cancer patients are now available where family members of cancer patients are discriminated against in society. One such example is that of the family members of Henrietta Lacks (Skloot, 2017). Here, the medical researchers uploaded their genotypic information online, and when a complaint was filed against this, the information was taken down by the researchers. However, by the time it was taken down, the data had already been downloaded and was subsequently published either in parts or completely. In some other cases, the genomic information was uploaded without getting permission from the relatives of the patient (Nyholt et al., 2009; Ritter, 2013).

This discrimination can be more severe from a forensic perspective. Relatives of criminals can face extreme discrimination because of the seriousness of the matter. The racial imbalance in NDIS/CODIS data is quite clear. In the United States population, blacks are around 13%; however, in the database, they contribute around 40% of entries. Similarly, Latinxs make up 22% of the United States population but contribute 22% to database entries. This situation becomes more severe when one includes data related to immigrant detainees (Wetsman, 2020). On examination of online genomic databases, one would find that the records are predominantly from the white or Caucasian population, which makes up 75% of total database entries. This percentage is projected to rise to 90% in the coming years (Kaiser, 2018). This racial disparity in DNA databases can be due to two factors: (a) Disparity in people who get arrested, and (b) disparity in resources to get out of it. Even in countries such as the United States, the financial divide between the white population and the population of other races is significant, so the white population is more likely to go through DNA sequencing to store their genomic data in a database (Klugman & Rodriguez, 2021). Based on this, crimes related to the white population have a significantly higher chance of getting solved using these genomic databases.

There is another facet to this problem. NDIS/CODIS was developed to store the data of criminals for their future identification, and this was acceptable. Eventually, many organizations uploaded the relevant data even when the person was arrested and later let go, or even when the person was arrested for minor crimes. Once the person's data is uploaded to the database, he/she is presumed to be guilty of the crime without proof in a court of law. Extending it to the relatives of the perpetrator might be a breach of the relative's privacy. In Maryland v. King, the US Supreme Court held that by committing the crime, the criminal surrenders their right to privacy (Field et al., 2017). The important question here is whether this also applies to the relatives of the criminal (Haimes, 2006). Do the relatives of the criminal give up their right to privacy just by being related to the criminal? Should the families of criminals be subjected to lifelong

surveillance and indirect lifelong surveillance? If not, then why should they be involved in criminal investigations? The family should not be subjected to any such discrimination, as every person is equal before the law, irrespective of their relationship with the criminal, and must be innocent until proven guilty. Genomic data of a person can reveal their phenotypic information, which can lead to additional discrimination based on race, gender, or any other medical or criminal features (Klugman & Rodriguez, 2021). The use of genomic data might be considered easy and convenient; however, considering the discriminatory possibility, it must be used with caution.

Bypassing codes of informed consent

The process of obtaining consent is a key element of medical and biomedical research. It makes the donor autonomous in choosing whether or not to donate their genomic data by informing them of the risks and benefits associated with the application (Hoeyer & Lynöe, 2006; Tindana et al., 2019). This autonomous choice has three features: (a) Adequate information; (b) voluntary or coerced choice; and (c) rationality in choice. In the case of IGG databases, donors should be clearly explained the information regarding consent in simple, clear, and complete words (Wickenheiser, 2019).

A biomedical biobank is developed to fight critical diseases and supplement medical knowledge. In contrast, the aim of using genomic databases for criminal investigations is to identify the perpetrators using pedigree or familial relationships (Machado & Silva, 2015). Therefore, consent in both settings should be explained differently. For example, a person might readily give his/her consent for medical research or to identify his/her relatives. The same person might not be willing to give consent if the information is to be used for a criminal investigation. Therefore, it is important to plan guidelines that are specific to the use of genomic data in criminal investigations. The process of consent, if efficient, can turn the risks of IGG into its strengths (Court, 2018). If neglected, it can severely affect the credibility of the IGG databases.

In a 2007 Nuffield Council study, it was observed that over 40% of donors gave unlimited consent for the usage of their genomic data for research purposes, which was later monetized, and then the consent could not be withdrawn (Court, 2018; Parry, 2008). There are some serious concerns associated with the consent information and how it is delivered to the donor. It is important that the donor is informed correctly of his rights, risks, and duties and that he understands this information. Many companies only inform donors that their genomic data will be used for medical or research purposes and not for investigation (Berkman et al., 2018). Some companies, such as GEDmatch and FamilyTreeDNA, allow the use of databases for investigative purposes. They only ask for broad consent rather than specific case-to-case consent, which is important in such cases. The settings of the consent can be changed at later stages as well (Ram & Roberts, 2019; Roberts & Hawkins, 2020). Other companies use undecipherable or incomplete information that the donor does not read or understand (Radin, 2012).

How the consent guidelines should be written is yet to be resolved. Some studies suggest that people know about the importance of consent while collecting the samples; however, they are divided on how specific the consent must be (Grady et al., 2015; Wendler, 2006). Some individuals believe that broad consent is acceptable, while others emphasize that more information must be provided to make an educated decision, and some others emphasize that the information provided should be in such a way that it is easily understandable (Madrigal, 2012; Samuel & Kennett, 2021).

Genetic surveillance

One of the many ways of misusing the genetic information available in IGG databases is to use genomic data for genetic surveillance (Zaretsky, 2021). Genetic surveillance includes tracing the individual and his family using their genomic information. Databases often have an overabundance of data from specific populations, and therefore, this information can be used to find individuals of a specific race, ethnicity, caste, or religion (Rosen, 2009). Professor Mnookin has called the process of familial searching a discriminator and adds that this discrimination is not just limited to race or class but extends to good relatives because of some bad ones (Epstein, 2009).

Via genealogical databases

Users upload their genotypes to the direct-to-customer databases, which are online, to search for their long-lost relatives. Once they upload their data, the database searches it to generate a report. This report may contain substantial information about one's putative relatives (Edge & Coop, 2020). Identical by state (IBS) genomic runs of (near) identical sequences shared among individuals can be thought of as a superset of true IBD segments. Edge and Coop discuss the methods by which the genetic privacy of the user can be compromised. They describe three methods for revealing the gene sequencing data in these online databases, where one can upload their own genotype (Edge & Coop, 2020). These methods are: (1) IBS tiling; (2) IBS probing; and (3) IBS baiting.

In IBS tiling, the genotype information shared between a target user in the database and each member of a set of comparison genomes is aggregated into potentially substantial information about the target's genotype (Fig. 26.5). For example, a person of Caucasian origin is likely to have some genetic similarity with other people of the same origin across the world. By comparing the target user's genotype with that of a random person of the same descent, one can gather some information. If we extend this method to more individuals of the same origin, we can gather even more information about the target individual. By combining all the gathered information, one can uncover much of the target's genome through the process of "tilling" of shared IBS with known genotypes. Compared to IBS segments, true IBD segments can reveal more information about shared genotypes. This is because of the higher probability

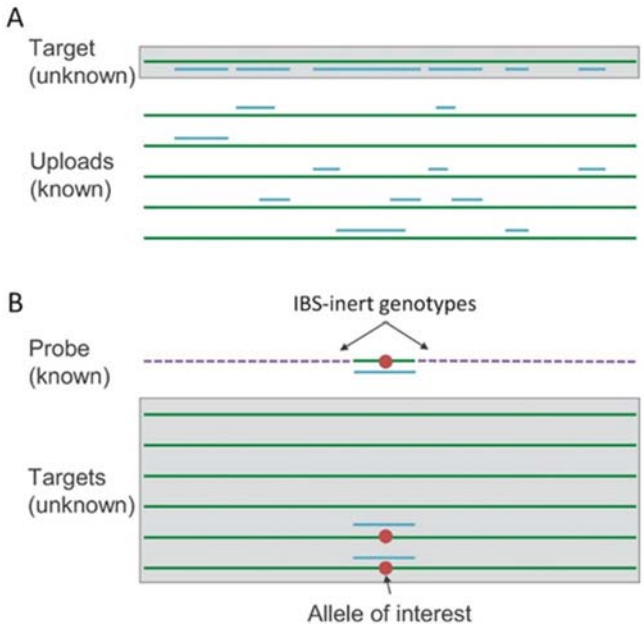


Figure 26.5 (A) IBS tiling and (B) IBS probing.

of untyped variants getting shared among individuals of similar origin. The distinction between two IBS, where IBS1 covers only one chromosome and IBS2 both chromosomes, is difficult in the absence of haplotype information. IBS tiling provides high population variation. It is difficult to differentiate between two IBS segments through this method using a haplotype. Therefore, the IBS tiling rates are varied, especially depending on the population under study. Genomic regions with low SNP density and low heterozygosity are prone to biases. This means that the recovery of minor alleles using IBS tiling is slightly less in comparison to the genomic population (Edge & Coop, 2020).

IBS probing is used to identify people of a specific genotype (Fig. 26.5). For example, if one wants to identify people having the risk alleles of, let's say, Parkinson's disease, even when there is no record of this information. All one has to do is prepare a haplotype with a specific genotype and search the public database. The database will search through it and suggest all the people who match this specific allele. For this, a chromosome that is unlikely to have long shared segments is prepared so that it is free of linkage disequilibrium. Long matching segments represent close relatives, and therefore, their presence reduces the sensitivity of the IBS probing method. The success of the probing method also depends on the frequency of the allele of interest in the given population. The rarer the allele of interest, the more likely the results are to be significant. If the allele of interest is common, there is a higher chance that they will share the same long haplotype around

the site of interest, and thus fewer probes will be required to find out the identities of the carriers in the given population (Edge & Coop, 2020).

IBS baiting uses incompatible homozygous sites, which are homozygous in pairs for one allele in one individual and homozygous in pairs for another allele in another individual. In this method, using heterozygous sites, IBS segments are constructed. Using this constructed IBS segment, the genotype of the target sequence is revealed. IBS baiting is of two types: (a) Single-site IBS baiting and (b) Parallel IBS baiting. In the case of single-site IBS baiting, a sequence at a single site is revealed, whereas in the case of parallel IBS baiting multiple sequences are revealed simultaneously (Fig. 26.6) (Edge & Coop, 2020).

Ethical issues

Forced collection of DNA samples from arrested individuals

It is common practice to collect an individual's DNA sample on arrest as a procedure, even when the crime has not been proven in a court of law. It is considered that the accused has forsaken his right to privacy. However, jumping to this conclusion can be very problematic. First, taking and analyzing DNA samples is part of a search, which

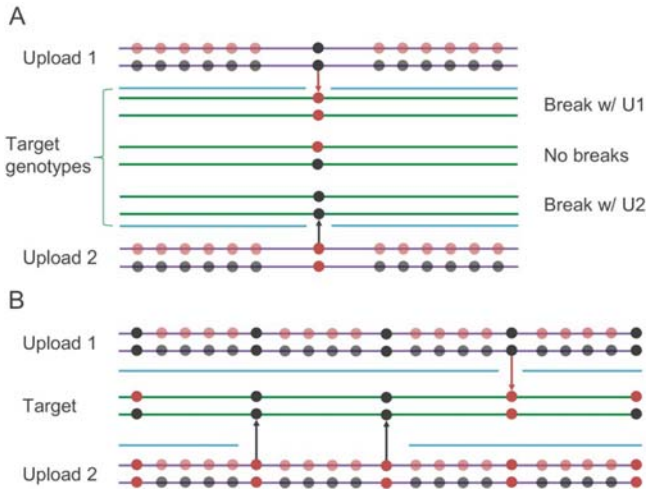


Figure 26.6 Schematics of the IBS baiting procedure (A) To perform IBS baiting at a single site, two uploads are required, each with runs of heterozygous genotypes flanking the key site. At the key site, the two uploaded datasets are homozygous for different alleles. The three possible target genotypes at the key site can each be determined by examining their IBS coverage with the uploads. If there is a break in IBS with either upload, then the target is homozygous for the allele not carried by the upload that shows the break in IBS (with the broken IBS segment shown as a cyan line). If there is no break in IBS with either upload, then the target is heterozygous at the key site. (B) Target genotypes at many key sites across the genome can be learned by comparison with two uploaded datasets, as long as key sites are spaced widely enough.

must not be started without probable cause and judicial authorization. Second, this mandatory DNA sample collection must not be considered a part of a special need, especially when the guilt has not been proven. Based on the first two features, the forceful collection of DNA samples must be unconstitutional (Rosen, 2009). DNA sampling must be considered a search because it includes public exposure, bodily intrusion, and information extracted. Even though the criminal has forsaken their right to privacy, no consent has been given by the relatives of the perpetrator to use their genomic data to incriminate their relatives (Court, 2018). This applies to all the databases storing the genomic data, whether for any medical research or for investigative purposes.

Lack of a law protecting privacy

Every country has some fundamental rights and laws protecting privacy, but laws protecting genetic privacy are few and rare. The debate on genetic privacy is ongoing across countries. Many states in the United States have implemented laws protecting genetic privacy, and some others are in the process of formulating such laws. Some European countries have also come forward to develop such laws. Despite such steps, there is a dire need for uniform laws against genetic privacy breaches (Hudson et al., 2008; Norrgard, 2008). Many laws, such as the Fourth Amendment of the United States Constitution, Section 20, and Section 21 of the Indian Constitution, provide protection against unauthorized searches, seizures, and sample collection (Baird & Avots, 2008; Khosla, 2012). This also includes the rights of the dead, which are often termed postmortem privacy. What if the deceased did not desire their genomic data to be used in the database? In such cases, the genomic data must be used with caution.

Case studies

Phoenix canal killings

The canal killing in the Phoenix area is an important case that emphasizes the use of GG to solve the mystery. In 1990, two women in the Phoenix area were murdered. Bryan Patrick Miller was linked using DNA findings to the deaths of Angela Brosso (November 1992) and Melanie Bernas (September 1993). This was the first-ever case that was solved using genetic genealogy in the year 2015.

Golden State serial killer or case of Joseph DeAngelo

In 2018, IGG was successfully used in the Investigation of the Golden State serial killer case to identify East Area Rapist Joseph James DeAngelo for committing 45 rapes/sexual assaults and 13 murders in California in the 1970s and 1980s. Investigators collected left-over cans as evidence from outside of DeAngelo's house, and eventually a piece of tissue from the trash was used to get the arrest warrant against DeAngelo. Investigators uploaded the DNA profile of the suspect collected from the crime scene

of the murder in 1980 to a genealogy GEDmatch genealogical database and traced out the familial tree of the suspect based on closed genetic matches (Mehar, 2021; Sarbak & Won, 2023).

Way forward

The use of genetic genealogy is only two decades old, as its first usage was reported in the year 2000. Its prime use was in the field of medical research, where it was being used to track the genotype sequences that were responsible for serious medical conditions such as cancer, etc. Starting with the Golden State killer case, the use of genetic genealogy in criminal investigations caught pace. Despite many analytical issues (errors), ethical issues, privacy issues, and other challenges, their use is increasing exponentially, as is evident by the ballooning growth of the genetic genealogy market. Table 26.2 enumerates key challenges and possible solutions that can make the IGG practice more reliable, credible, and authentic. The first requirement is to set up guidelines for the use of IGG and protocols for the methods, techniques, and application of IGG. The second is to simplify while explaining in detail the implications, risks, threats, and applications of the consent before sampling and during its use in the criminal investigation. Another important requirement is to make sure that volunteers do not just accept the terms and conditions but rather understand the consent

Table 26.2 Challenges and solutions to address challenges related to IGG.

S.No	Challenges	Possible solution
1.	Technological and phase errors	Use a haplotype score incorporating the phase and genotype error rates
2.	Presence of DNA mixtures on the crime scene	Separate profiles of the contributors
3.	Degraded or low-quality samples	Pretreatment to remove contamination
4.	Need for standards	Formulation of standards
5.	False identification	Restrict usage of IGG for finding distant relatives
6.	Use of unproven or invalid testing methods	Standardize the methods and protocols
7.	Biospecimen and hacking threats	Measures to secure the cloud systems
8.	Genomic privacy	Regulation to protect genomic privacy
9.	Informed consent	Formulate guidelines for obtaining consent
10.	Genetic surveillance	Laws against prejudicial and unauthorized genetic surveillance
11.	Ethical issues (forced collection of DNA samples from arrested individuals and lack of laws protecting privacy)	Formulate guidelines and laws to address these issues

information properly. The third is to clearly explain the process of the use of genomic data for criminal investigation and how it eventually affects the final verdict in court. Just the presence of DNA at a crime scene does not prove that the person has committed the crime. Therefore, the information provided by IGG must not be taken as absolute proof of criminal activity. Fourth, guidelines against the use of IGG for prejudicial familial searching, genetic surveillance, and other types of discrimination must be formulated and enforced. There must be a mechanism of grievance in cases of genetic privacy breaches. Fifth, in accordance with these guidelines, there must be laws pertaining to genetic privacy. IGG is nascent and is still evolving, and it is mandatory for IGG practitioners, investigating agencies, volunteers, and legislative bodies to work in accordance to bring out the best of IGG.

Conclusions

Genetic genealogy has come a long way since its inception and has the potential for even higher growth. It has immense application in the field of medical research. Its application in the field of criminal investigation is very recent and must be used with great caution. Ethical and privacy concerns, prejudicial familial searches, and genetic surveillance are some of the key challenges of IGG. It is important that IGG practitioners come together to address these issues and formulate the necessary guidelines for the unbiased use of IGG.

References

Aldhous, P. (2020). *A security breach exposed more than one million DNA profiles on a major genealogy database.*

Apathy, N. C., Menser, T., Keeran, L. M., Ford, E. W., Harle, C. A., & Huerta, T. R. (2018). Trends and gaps in awareness of direct-to-consumer genetic tests from 2007 to 2014. *American Journal of Preventive Medicine, 54*, 806–813.

Baird, C. F., & Avots, K. (2008). Fourth amendment. *Texas Tech Law Review, 41*, 961.

Ball, C. A., Barber, M. J., Byrnes, J., Carbonetto, P., Chahine, K. G., Curtis, R. E., Granka, J. M., Han, E., Hong, E. L., Kermany, A. R., Myres, N. M., Noto, K., Qi, J., Rand, K., Wang, Y., & Willmore, L. (2016). Ancestry DNA matching white paper. *Ancestry Com*, 1–46.

Berkman, B. E., Miller, W. K., & Grady, C. (2018). Is it ethical to use genealogy data to solve crimes? *Annals of Internal Medicine, 169*, 333–334.

Bleka, Ø., Storvik, G., & Gill, P. (2016). EuroForMix: An open source software based on a continuous model to evaluate STR DNA profiles from a mixture of contributors with artefacts. *Forensic Science International: Genetics, 21*, 35–44. https://doi.org/10.1016/j.fsigen.2015.11.008

Bradford, L., Heal, J., Anderson, J., Faragher, N., Duval, K., & Lalonde, S. (2011). Disaster victim investigation recommendations from two simulated mass disaster scenarios utilized for user acceptance testing CODIS 6.0. *Forensic Science International: Genetics, 5*, 291–296. https://doi.org/10.1016/j.fsigen.2010.05.005

Bright, J.-A., Taylor, D., McGovern, C., Cooper, S., Russell, L., Abarno, D., & Buckleton, J. (2016). Developmental validation of STRmix™, expert software for the interpretation of forensic DNA profiles. *Forensic Science International: Genetics, 23*, 226–239. https://doi.org/10.1016/j.fsigen.2016.05.007

Browning, S. R., & Browning, B. L. (2007). Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *The American Journal of Human Genetics, 81*, 1084–1097. https://doi.org/10.1086/521987

Browning, B. L., & Browning, S. R. (2013). Detecting identity by descent and estimating genotype error rates in sequence data. *The American Journal of Human Genetics, 93*, 840–851. https://doi.org/10.1016/j.ajhg.2013.09.014

Callaghan, T. F. (2019). Responsible genetic genealogy. *Science, 366*(80), 155.

Carroll, N. M., Blum-Barnett, E., Madrid, S. D., Jonas, C., Janes, K., Alvarado, M., Bedoy, R., Paolino, N., Aziz, V., McGlynn, E. A., & Burnett-Hartman, A. N. (2020). Demographic differences in the utilization of clinical and direct-to-consumer genetic testing. *Journal of Genetic Counseling, 29*, 634–643.

Court, D. S. (2018). Forensic genealogy: Some serious concerns. *Forensic Science International: Genetics, 36*, 203–204.

Davawala, A., Stock, A., Spiden, M., Daniel, R., McBain, J., & Hartman, D. (2022). Forensic genetic genealogy using microarrays for the identification of human remains: The need for good quality samples – a pilot study. *Forensic Science International, 334*, 111242. https://doi.org/10.1016/j.forsciint.2022.111242

Deb, F. (2022). *Role of population genetics and genetic genealogy in forensic*. SIFS INDIA.

Dimitromanolakis, A., Paterson, A. D., & Sun, L. (2019). Fast and accurate shared segment detection and relatedness estimation in un-phased genetic data via TRUFFLE. *The American Journal of Human Genetics, 105*, 78–88. https://doi.org/10.1016/j.ajhg.2019.05.007

Direct-to-consumer genetic testing market, by test type (carrier testing and predictive testing), by technology (targeted analysis and whole genome sequencing), and by region forecast to 2030, 2022. EMERGEN Research. Report ID: ER_0020. https://www.emergenresearch.com/industry-report/direct-to-consumer-genetic-testing-market. accessed on 05-08-2023.

Duarte Castelão, C. (2017). *Next gen sequencing challenges*.

Durand, E. Y., Eriksson, N., & McLean, C. Y. (2014). Reducing pervasive false-positive identical-by-descent segments detected by large-scale pedigree Analysis. *Molecular Biology and Evolution, 31*, 2212–2222. https://doi.org/10.1093/molbev/msu151

Edge, M. D., & Coop, G. (2020). Attacks on genetic privacy via uploads to genealogical databases. *Elife, 9*, e51810.

Epstein, J. (2009). "Genetic surveillance"-the bogeyman response to familial DNA investigations. *Journal of Law, Technology and Policy*, 141.

Epstein, M. P., Duren, W. L., & Boehnke, M. (2000). Improved inference of relationship for pairs of individuals. *The American Journal of Human Genetics, 67*, 1219–1231. https://doi.org/10.1016/S0002-9297(07)62952-8

Field, M. B., Seera, S., Nguyen, C., & Debus-Sherrill, S. (2017). *Study of familial DNA searching policies and practices: Case study brief series*. ICF.

Glynn, C. L. (2022). Bridging disciplines to form a new one: The emergence of forensic genetic genealogy. *Genes, 13*, 1381.

Grady, C., Eckstein, L., Berkman, B., Brock, D., Cook-Deegan, R., Fullerton, S. M., Greely, H., Hansson, M. G., Hull, S., Kim, S., Lo, B., Pentz, R., Rodriguez, L., Weil, C., Wilfond, S., & Wendler, D. (2015). Broad consent for research with biological samples: Workshop conclusions. *The American Journal of Bioethics, 15*, 34–42.

Greytak, E. M., Moore, C., & Armentrout, S. L. (2019). Genetic genealogy for cold case and active investigations. *Forensic Science International, 299*, 103–113.

Guerrini, C. J., Wickenheiser, R. A., Bettinger, B., McGuire, A. L., & Fullerton, S. M. (2021). Four misconceptions about investigative genetic genealogy. *The Journal of Law and the Biosciences, 8*, lsab001.

Gurney, D., Press, M., Moore, C., Rolnick, C. I., Hochreiter, A., & Bossert, B. L. (2022). The need for standards and certification for investigative genetic genealogy, and a notice of action. *Forensic Science International, 341*, 111495. https://doi.org/10.1016/j.forsciint.2022.111495

Haimes, E. (2006). Social and ethical issues in the use of familial searching in forensic investigations: Insights from family and kinship studies. *Journal of Law Medicine & Ethics, 34*, 263–276.

Haned, H. (2011). Forensim: An open-source initiative for the evaluation of statistical methods in forensic genetics. *Forensic Science International: Genetics, 5*, 265–268. https://doi.org/10.1016/j.fsigen.2010.03.017

Henn, B. M., Hon, L., Macpherson, J. M., Eriksson, N., Saxonov, S., Pe'er, I., & Mountain, J. L. (2012). Cryptic distant relatives are common in both isolated and cosmopolitan genetic samples. *PLoS One, 7*. https://doi.org/10.1371/journal.pone.0034267

Hoeyer, K., & Lynöe, N. (2006). Motivating donors to genetic research? Anthropological reasons to rethink the role of informed consent. *Medicine, Health Care and Philosophy, 9*, 13–23.

Hudson, K. L., Holohan, M. K., & Collins, F. S. (2008). Keeping pace with the times—the genetic information nondiscrimination act of 2008. *New England Journal of Medicine, 358*, 2661–2663.

Kaiser, J. (2018). We will find you: DNA search used to nab Golden state killer can home in on about 60% of white Americans. *Science*, (80).

Khosla, M. (2012). *The Indian constitution*. Oxford University Press.

Kling, D. (2019). On the use of dense sets of SNP markers and their potential in relationship inference. *Forensic Science International: Genetics, 39*, 19–31. https://doi.org/10.1016/j.fsigen.2018.11.022

Kling, D., Phillips, C., Kennett, D., & Tillmar, A. (2021). Investigative genetic genealogy: Current methods, knowledge and practice. *Forensic Science International: Genetics, 52*, 102474. https://doi.org/10.1016/j.fsigen.2021.102474

Klugman, C. M., & Rodriguez, H. F. (2021). Ethics of familial genetic genealogy: Solving crimes at the cost of privacy. *DePaul Journal of Health Care Law, 22*, 67.

Lindor, N. M., Johnson, K. J., McCormick, J. B., Klee, E. W., Ferber, M. J., & Farrugia, G. (2013). Preserving personal autonomy in a genomic testing era. *Genetics in Medicine: Official Journal of the American College of Medical Genetics, 15*, 408–409. https://doi.org/10.1038/gim.2013.24

Machado, H., & Silva, S. (2015). Public participation in genetic databases: Crossing the boundaries between biobanks and forensic DNA databases through the principle of solidarity. *Journal of Medical Ethics, 41*, 820–824.

Madrigal, A. C. (2012). Reading the privacy policies you encounter in a year would take 76 work days. *The Atlantic, 1*.

McDermott, M. (2020). Investigative genetic genealogy: How does it work? *National Genealogical Society*. https://www.ishinews.com/investigative-genetic-genealogy-how-does-it-work/, accessed on 05-08-2023.

Mehar, P. A. (2021). *Forensic genetic genealogy: An investigative aid*.

Naveed, M., Ayday, E., Clayton, E. W., Fellay, J., Gunter, C. A., Hubaux, J.-P., Malin, B. A., & Wang, X. (2015). Privacy in the genomic era. *ACM Computing Surveys, 48*. https://doi.org/10.1145/2767007

Norrgard, K. (2008). Protecting your genetic identity: GINA and HIPAA. *Nature Educationa, 1*, 21.

Nyholt, D. R., Yu, C.-E., & Visscher, P. M. (2009). On Jim Watson's APOE status: Genetic information is hard to hide. *European Journal of Human Genetics, 17*, 147–149.

Parry, B. (2008). The forensic use of bioinformation: A review of responses to the Nuffield report, published in BioSocieties, volume 3(1). *BioSocieties, 3*, 217–222. https://doi.org/10.1017/S1745855208006121

Phillips, C. (2018). The Golden State Killer investigation and the nascent field of forensic genealogy. *Forensic Science International: Genetics, 36*, 186–188. https://doi.org/10.1016/j.fsigen.2018.07.010

Radin, M. J. (2012). *The fine print, vanishing rights, and the rule of law*.

Ram, N., Murphy, E. E., & Suter, S. M. (2021). Regulating forensic genetic genealogy. *Science, 373*(80), 1444–1446.

Ram, N., & Roberts, J. L. (2019). Forensic genealogy and the power of defaults. *Nature Biotechnology, 37*, 707–708.

Ritter, M. (2013). *Henrietta Lacks' family, feds reach settlement on use of DNA info*.

Roberts, J. L., & Hawkins, J. (2020). When health tech companies change their terms of service. *Science, 367*(80), 745–746.

Rosen, J. (2009). Genetic surveillance for all. *Slate Magazine, 17*.

Salloum, R. G., George, T. J., Silver, N., Markham, M.-J., Hall, J. M., Guo, Y., Bian, J., & Shenkman, E. A. (2018). Rural-urban and racial-ethnic differences in awareness of direct-to-consumer genetic testing. *BMC Public Health, 18*, 1–6.

Samuel, G., & Kennett, D. (2020). The impact of investigative genetic genealogy: Perceptions of UK professional and public stakeholders. *Forensic Science International: Genetics, 48*, 102366. https://doi.org/10.1016/j.fsigen.2020.102366

Samuel, G., & Kennett, D. (2021). Problematizing consent: Searching genetic genealogy databases for law enforcement purposes. *New Genetics & Society, 40*, 284–304.

Sarbak, M., & Won, K. (2023). *Idaho student murders: What is genetic genealogy, a tool reportedly used to help capture the suspect?*.

Saunders, I. W., Brohede, J., & Hannan, G. N. (2007). Estimating genotyping error rates from Mendelian errors in SNP array genotypes and their impact on inference. *Genomics, 90*, 291–296. https://doi.org/10.1016/j.ygeno.2007.05.011

Scudder, N., Daniel, R., Raymond, J., & Sears, A. (2020). Operationalising forensic genetic genealogy in an Australian context. *Forensic Science International, 316*, 110543.

Skloot, R. (2017). *The immortal life of Henrietta Lacks*. Broadway Paperbacks.

Swain, R., & Subodh, K. (2023). *Direct-to-Consumer (DTC) genetic testing market size by test type (carrier testing, predictive testing, ancestry and relationship testing, nutrigenomic testing, skincare testing), by technology (targeted analysis, single nucleotide polymorphism (SNP) chips*.

Thomson, J., Clayton, T., Cleary, J., Gleeson, M., Kennett, D., Leonard, M., & Rutherford, D. (2020). An empirical investigation into the effectiveness of genetic genealogy to identify individuals in the UK. *Forensic Science International: Genetics, 46*, 102263.

Tillmar, A., Fagerholm, S. A., Staaf, J., Sjölund, P., & Ansell, R. (2021). Getting the conclusive lead with investigative genetic genealogy – a successful case study of a 16 year old double murder in Sweden. *Forensic Science International: Genetics, 53*, 102525. https://doi.org/10.1016/j.fsigen.2021.102525

Tindana, P., Molyneux, S., Bull, S., & Parker, M. (2019). 'It is an entrustment': Broad consent for genomic research and biobanks in sub-Saharan Africa. *Developing World Bioethics, 19*, 9–17. https://doi.org/10.1111/dewb.12178

Wendler, D. (2006). One-time general consent for research on biological samples: Is it compatible with the health insurance portability and accountability act? *Archives of Internal Medicine, 166*, 1449–1452.

Wetsman, N. (2020). DNA from detained immigrants will change the nature of the FBI's genetic database. *Verge*.

Wickenheiser, R. A. (2019). Forensic genealogy, bioethics and the Golden state killer case. *Forensic Science International: Synergy, 1*, 114–125.

Zaretsky, A. (2021). DNA collection in immigration custody and the threat of genetic surveillance. *California Law Review, 109*, 317.

CHAPTER 27

The role of artificial intelligence and machine learning in NGS

Khaled Ouanes
Health Informatics Department, Health Sciences College, Saudi Electronic University, Dammam, Saudi Arabia

What is next-generation sequencing?

Next-generation sequencing (NGS) is a relatively novel technology that enables researchers to sequence large amounts of genetic material quickly and cost-effectively. NGS is based on the process of sequencing short fragments of DNA or RNA, which are then assembled into a complete sequence of the genome.

NGS techniques use modified nucleotides, polymerases for synthesis, and fluorescence detection, which is the same idea behind Sanger sequencing. Although the DNA template must be clonally amplified for some NGS platforms, such as Illumina, Life Technologies SOLiD, Ion Torrent Personal Genome Machine, and Roche 454 systems, preamplification is not required for the more sensitive HeliScope and Pacific Biosciences single-molecule real-time systems (Buermans et al., 2014; Stranneheim et al., 2012). Different NGS platforms use various chemistries for the preparation of libraries and sequencing (Hodzic et al., 2017).

We can confirm that NGS enables scientists and investigators to sequence large amounts of genetic material contained in different nucleic acids at once. This would otherwise be virtually impossible or impractical to achieve with other sequencing technologies. Indeed, researchers have been able to sequence the entire human genome in a single experiment and make discoveries that would have been impossible in the past.

Virtually all downstream analysis and interpretation processes rely on accurate variant calling, which is a crucial step in the processing of NGS data. The software tools, technologies, and methods for finding sequence variants have advanced significantly during the past few years, clearly paralleling the advancement of NGS technologies (Koboldt, 2020).

Next-generation sequencing for forensics

NGS, despite being a relatively new method, has been quite useful for forensic sciences and investigations in the past few years. Indeed, it can be used for sequencing whole genomes or portions of genomes with a high degree of accuracy (Zhang et al., 2019).

Next Generation Sequencing (NGS) Technology in DNA Analysis
ISBN 978-0-323-99144-5, https://doi.org/10.1016/B978-0-323-99144-5.00002-0
© 2024 Elsevier Inc.
All rights reserved.

It has been applied to forensic DNA analysis, allowing examiners to generate data that spans the human genome and addresses a wider range of questions in a single, targeted assay (Zhang et al., 2019). NGS-generated short tandem repeat (STR) calls are completely consistent and compatible with current database formats. Additionally, NGS technology can be used to identify organ- and developmental stage-specific expression as well as miRNA sequences, and some single nucleotide polymorphisms (SNPs) may be used to predict the bio-geographical ancestry of unknown crime scene samples (Carratto et al., 2022).

We can clearly see that the rapid emergence of NGS techniques, also sometimes known as massively parallel sequencing methodologies, has revolutionized many genomic fields, including forensics. In 2005, Roche 454 introduced the first high-throughput sequencing platform based on the pyrosequencing technique.

Since then, NGS methodology has undergone significant enhancements and improvements, with decreased costs and analysis time and increased sensitivity. Robust tools and instruments have been developed and are currently commercially available. Although standard DNA analyses in forensic laboratories involve STR profiling using capillary electrophoresis (CE), this methodology has certain limitations, including its inability to distinguish monozygotic twins or work with low quantities of samples. Low-quality samples can also hinder the accuracy of the analysis (Carratto et al., 2022). Additionally, the CE methodology presents a lower multiplex capability. Thus, new approaches are necessary and even paramount to being able to extract more information from DNA to aid forensic casework.

NGS presents several advantages and improvements over conventional CE analyses, including the capability of analyzing thousands of markers of various types at once, providing different information in a single analysis. NGS can also better handle the low-quantity and degraded samples that are common in forensic investigations. Therefore, NGS is deemed to be a powerful tool for deeper DNA analysis in forensic genetics, leading to its steady introduction and widespread use as a standard method in forensic laboratories.

In a recent survey from 2019 of 105 European forensic laboratories, 46% of participants already owned NGS equipment and consumables, while 26.7% planned to acquire them within the next 2 years (Gross et al., 2021). Clearly, NGS analysis can empower the study of polymorphisms, opening new horizons and opportunities for forensic genetics and molecular biology.

Moreover, in situations like mass catastrophes, man-made disasters, terrorist attacks, or other events where forensic specimens and samples are contaminated and degraded, NGS technologies will be essential for human DNA analysis and typing. NGS will make it feasible to analyze the usual autosomal DNA (STRs and SNPs), mitochondrial DNA, and X and Y chromosomal markers all at once and yield robust results (Alvarez-Cubero et al., 2017).

What is artificial intelligence? Why is it useful for forensics?

Artificial intelligence (AI) is revolutionizing many industries, including the fields of forensic omics, genomics, and NGS. AI refers to the development of computerized systems and solutions that can perform tasks that normally require human intelligence, such as perception, reasoning, learning, and decision-making. The field of AI encompasses a range of techniques, including machine learning (ML), natural language processing, computer vision, and robotics, among others. AI has the potential and capability to revolutionize and drastically transform many areas of different fields and industries, from healthcare and finance to transportation and education.

One of the earliest and most influential definitions of AI was proposed by John McCarthy, who described it as "the science and engineering of making intelligent machines"(McCarthy, 1956). Since then, the field has evolved and expanded significantly, with many researchers and practitioners offering their own definitions and perspectives. For instance, Russell and Norvig (2010) define AI as "the study of agents that receive perceptions from the environment and take actions that affect that environment."

While AI has made significant progress in recent years, it is still a rapidly evolving field with many challenges and opportunities. Some of the key research areas in AI today include deep learning, ML, reinforcement learning, explainable AI, and the ethical and societal implications of AI. This multifaceted and intricate field is witnessing novel developments and ceaselessly accelerating breakthroughs occurring frequently. In recent years, there have been several notable advancements in AI, including in the areas of natural language processing, computer vision, ML, and deep learning. All these areas can be very useful for biomedical sciences, forensics, and genomics.

Despite these impressive developments, AI still faces many challenges, particularly in areas such as explainability, interpretability, fairness, and ethics. Explainability is especially crucial for users interacting with AI to recognize and fathom the AI's interpretations, conclusions, and recommendations. As AI continues to evolve, it will be important for researchers and practitioners to address these challenges and ensure that AI is used in a responsible and beneficial way. This is even more central for forensic studies and casework, especially from an ethical point of view.

AI is increasingly being used in biomedical sciences and forensics to improve our understanding of diseases and enhance forensic investigations. In biomedical sciences, AI has been applied to various tasks, including drug discovery, medical imaging analysis, and precision medicine (Gómez-Ríos et al., 2021; Krittanawong et al., 2021). AI has also been used in precision medicine to develop personalized treatment plans based on an individual's genetic and medical history (Wang & Preininger, 2019).

In forensics, AI has been used to analyze various types of evidence, including DNA, fingerprints, and surveillance footage. For example, AI can be used to analyze DNA sequences to identify suspects or determine familial relationships. Facial recognition

software, which uses AI algorithms to match images to a database of known individuals, has also been used in forensic investigations (Jain et al., 2011). AI can also be used to analyze surveillance footage to identify suspects or track their movements.

Most importantly, AI has been used in forensic genomics for various purposes, such as identifying suspects, predicting the physical traits of suspects from DNA samples, and identifying the tissue source of forensic samples. The use of AI in forensic genomics can be traced back to the 1990s, when early primitive ML algorithms were first applied to DNA sequencing data (Baldi & Brunak, 2001).

In recent years, AI has been used for more advanced applications in forensic genomics. For example, in 2019, a team of researchers used a deep learning algorithm to predict the physical appearance of suspects from DNA samples, including eye, skin, and hair color (Chaitanya et al., 2019). In another study, researchers used AI to identify the tissue source of forensic samples based on gene expression patterns, which could aid in the identification of missing persons or unidentified human remains (Korsak et al., 2019).

Despite the potential benefits of using AI in forensic genomics, there are also concerns about the reliability of AI algorithms, particularly with regard to bias and accuracy. In addition, there are ethical considerations regarding the use of AI in forensic investigations, particularly with regard to the privacy and potential misuse of data (Bennett, 2020).

Considering all these points, we can comprehend that AI has a long history of use in forensic genomics, and recent advances in deep learning algorithms have enabled more advanced applications in this field. However, it is important to carefully consider the moral, ethical, and societal implications of AI in forensic investigations and to ensure that its use is in line with legal and ethical requirements and standards.

In addition, despite acknowledging the potential benefits of using AI in biomedical sciences and forensics, there are also concerns about privacy, bias, and the reliability of AI algorithms. It will be important for researchers, investigators, and practitioners to address these concerns and ensure that AI is used in a responsible and ethical way.

What is machine learning?

ML is a subfield of AI that enables computer systems to routinely and automatically learn and improve from experience without being explicitly programmed. In biomedical sciences, forensics, and genomics, ML has the potential to revolutionize the field by enabling faster and more accurate analysis of large amounts of generated data.

In biomedical sciences and forensics, ML has been already used for tasks such as medical image analysis, diagnosis of diseases, and drug discovery. For example, ML algorithms have been applied to medical imaging data to automatically detect abnormalities and assist with disease diagnosis (Litjens et al., 2017). ML has also been used to predict patient outcomes and develop personalized treatment plans based on a patient's genetic and medical history (Miotto et al., 2017).

In forensics, ML has been learning how to analyze various types of evidence and data, including DNA, fingerprints, and surveillance footage. For example, ML algorithms can be used to analyze DNA sequences to identify suspects or determine familial relationships (Li et al., 2019). ML can also be used to analyze patterns in surveillance footage to identify suspects or track their movements.

Despite the potential benefits of using ML in biomedical sciences, forensics, and genomics, there are also concerns similar to the ones raised about the overall use of AI in these fields. Indeed, matters like bioethics, privacy, bias, and the reliability of ML algorithms should be addressed. It will be important for researchers and experts to focus on these concerns, propose solutions to them, and ensure that ML is used in a responsible and ethical way.

How do artificial intelligence and machine learning help with NGS?

NGS generates vast amounts of data, making it challenging to analyze and interpret the results precisely. AI can help solve this problem by providing advanced algorithms and ML techniques to process and analyze NGS data quickly and accurately. Indeed, AI and ML are already used to improve the accuracy, reliability, efficiency, and speed of NGS. AI can be used to identify patterns in the data generated by NGS, which can help investigators and researchers interpret the results more quickly and accurately.

Another primary application of AI in NGS is data preprocessing, which involves quality control, adapter trimming, and filtering of low-quality reads. AI algorithms can automatically detect and remove sequencing errors and low-quality reads, which can significantly improve the accuracy and quality of the results. This is to say that AI can be used to identify and, consequently, flag many types of errors or abnormalities in the datasets, which can drastically reduce the risk of inaccuracies and mistakes in the produced final results. This will augment very clearly the power, robustness, and consistency of NGS technologies (Behjati & Tarpey, 2013; Kiran & Akhade, 2021; Sivakumar et al., 2021; Xu et al., 2021; Zhang et al., 2011).

AI and ML are also being used to improve data analysis in NGS. AI-based systems and ML algorithms can be used to identify patterns, variances, and anomalies, such as rare genetic variants or gene expression patterns, in large datasets generated by NGS. This will help researchers and investigators comprehend and interpret the results and identify important insights. ML algorithms can also be used to predict how certain patterns can be affected by certain elements, like DNA degradation. In addition, AI and ML can be used to detect genetic mutations, genetic variability, or diseases, which could lead to the development of new human identification tests.

Another important aspect is that AI and ML can be used to transform NGS methodologies and techniques, enabling researchers to generate more reliable results in even shorter amounts of time. Furthermore, ML algorithms can be used to build mathematical

models and simulations that can calculate and foresee the expected outcomes of any NGS analysis based on input training data. A very similar prediction approach was successfully developed relying on a deep learning model has been created for predicting very accurately NGS results (Zhang et al., 2021). These tools can be used to verify and validate NGS experimental results and outcomes.

Given the recognized influence of nucleic acid sequences on form and function and the exponential number of possible sequences of given lengths, the integration of generalizable standardized domain knowledge into deep learning architectures will be a noteworthy facilitator for predicting behaviors for nucleic acids in general (Zhang et al., 2021).

Challenges of utilizing AI and ML in NGS

While it is obvious that AI and ML can be used to improve the precision, robustness, and speed of NGS, there are also some challenges associated with using these technologies. These include the need for high-quality data, the complexity of the algorithms used, and the cost of implementation. Another major challenge is the lack of standardization in NGS data. Different sequencing platforms compounded with different protocols can generate data of varying quality, making it challenging to compare and analyze data from different sources and rendering the repeatability of the analysis very complicated.

Another challenge is the need for large amounts of high-quality data to train ML algorithms. Access to such data can be limited, particularly for rare diseases or disorders. In addition, the data generated by NGS can be complex and difficult to input and interpret, which can limit the effectiveness of AI and ML algorithms. Furthermore, many of the AI and ML algorithms used in NGS are still in the early stages of development, so it may be some time before they are completely ready for widespread use. However, with the current rate of innovation and development of new techniques, algorithms, and AI-based systems and tools, this latency time can be shorter than expected.

Conclusion

NGS is doubtlessly a revolutionary technology that has enabled researchers to sequence large amounts of genetic material and nucleic acids from different types of samples and tissues quickly and cost-effectively. AI and ML have the potential to augment and enhance NGS methods and techniques and to uncover new insights from the data they generate. However, there are also some challenges associated with utilizing AI and ML in NGS, including the need for high-quality data and the complexity of the procedures and systems used. With further research and development, AI and ML technologies could be used to greatly improve the accuracy and speed of NGS and open new

possibilities in the field of forensics, like they are doing for other fields. We can conclude that the application of AI and ML in NGS is still in its early stages, and there are many challenges to overcome. However, with continued research and development, AI can help unlock the full potential of NGS for forensics, personalized medicine, genetic analyses, and many other areas and fields.

References

Alvarez-Cubero, M. J., Saiz, M., Martínez-García, B., Sayalero, S. M., Entrala, C., Lorente, J. A., & Martinez-Gonzalez, L. J. (2017). Next generation sequencing: An application in forensic sciences? *Annals of Human Biology, 44*(7), 581–592. https://doi.org/10.1080/03014460.2017.1375155

Baldi, P., & Brunak, S. (2001). *The machine learning approach to protein structure*. MIT Press.

Behjati, S., & Tarpey, P. S. (2013). What is next generation sequencing? *Archives of Disease in Childhood-Education & Practice Edition, 98*(6), 236–238.

Bennett, D. (2020). Ethical considerations for the use of artificial intelligence in forensic investigations. *Journal of Forensic Sciences, 65*(1), 245–249.

Buermans, H. P., & den Dunnen, J. T. (2014). Next generation sequencing technology: Advances and applications. *Biochimica et Biophysica Acta, 1842*, 1932–1941. https://doi.org/10.1016/j.bbadis.2014.06.015

Carratto, T. M. T., Moraes, V. M. S., Recalde, T. S. F., Oliveira, M. L. G. D., & Teixeira Mendes-Junior, C. (2022). Applications of massively parallel sequencing in forensic genetics. *Genetics and Molecular Biology, 45*(3 Suppl. 1). https://doi.org/10.1590/1678-4685-gmb-2022-0077

Chaitanya, L., Breslin, K., Kerber, R. A., & Mitalipov, S. (2019). Predicting facial phenotypes of single individuals using whole-genome sequencing. *Frontiers in Genetics, 10*, 1148.

Gómez-Ríos, D., Gallo-Fernández, M., Rojas-Morales, P., & Ríos-García, J. (2021). Artificial intelligence and drug discovery: Future perspectives. *Expert Opinion on Drug Discovery, 16*(11), 1215–1228.

Gross, T. E., Fleckhaus, J., & Schneider, P. M. (2021). Progress in the implementation of massively parallel sequencing for forensic genetics: Results of a European-wide survey among professional users. *International Journal of Legal Medicine, 135*, 1425–1432.

Hodzic, J., Gurbeta, L., Omanovic-Miklicanin, E., & Badnjevic, A. (2017). Overview of next-generation sequencing platforms used in published draft plant genomes in light of genotypization of immortelle plant (*Helichrysium arenarium*). *Medical Archives, 71*(4), 288–292. https://doi.org/10.5455/medarh.2017.71.288-292

Jain, A. K., Klare, B., & Park, U. (2011). Face recognition: Some challenges in forensics. In *2011 IEEE International Conference on Automatic Face & Gesture Recognition (FG)* (pp. 726–733). Santa Barbara, CA, USA. https://doi.org/10.1109/FG.2011.5771338

Kiran, M., & Akhade, S. A. (2021). Artificial intelligence and machine learning in next-generation sequencing: Preprocessing, alignment, and variant calling. In *Artificial intelligence in healthcare* (pp. 263–292). Springer.

Koboldt, D. C. (2020). Best practices for variant calling in clinical sequencing. *Genome Medicine, 12*(1). https://doi.org/10.1186/s13073-020-00791-w

Korsak, A., Augustin, R., & Ballantyne, J. (2019). Forensic gene expression profiling from human organs and body fluids using a portable microfluidic system. *Forensic Science International: Genetics, 43*, 102159.

Krittanawong, C., Zhang, H., Wang, Z., Aydar, M., & Kitai, T. (2021). Artificial intelligence in precision cardiovascular medicine. *Journal of the American College of Cardiology, 77*(25), 3210–3225.

Li, C., Jiang, X., Zhao, H., & Yan, J. (2019). Forensic DNA phenotyping through machine learning: An empirical study. *Forensic Science International: Genetics, 43*.

Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., van der Laak, J. A. W. M., van Ginneken, B., & Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. *Medical Image Analysis, 42*, 60–88.

McCarthy, J. (1956). *Proposal for the dartmouth summer research project on artificial intelligence*.
Miotto, R., Wang, F., Wang, S., Jiang, X., & Dudley, J. T. (2017). Deep learning for healthcare: Review, opportunities and challenges. *Briefings in Bioinformatics, 19*(6), 1236–1246.
Russell, S. J., & Norvig, P. (2010). *Artificial intelligence: A modern approach*. Prentice Hall.
Sivakumar, S., Kumar, V. R., & Kumar, A. (2021). Next-generation sequencing: From bench to bedside. In *Principles and practice of clinical research* (pp. 39–57). Springer.
Stranneheim, H., & Lundeberg, J. (2012). Stepping stones in DNA sequencing. *Biotechnology Journal, 7*, 1063–1073. https://doi.org/10.1002/biot.201200153
Wang, F., & Preininger, A. (2019). AI in health: State of the art, challenges, and future directions. *Yearbook of Medical Informatics, 28*(1), 16–26. https://doi.org/10.1055/s-0039-1677908
Xu, B., Wu, Y., Xie, J., Zhang, R., & Yin, J. (2021). A comprehensive review of artificial intelligence applications in next-generation sequencing. *Journal of Molecular Biology, 433*(17), 2923–2944. https://doi.org/10.1016/j.jmb.2021.10.039
Zhang, J., Chiodini, R., Badr, A., & Zhang, G. (2011). The impact of next-generation sequencing on genomics. *Journal of Genetics and Genomics, 38*(3), 95–109. https://doi.org/10.1016/j.jgg.2011.02.003
Zhang, Y., & Wang, C. (2019). Application of next-generation sequencing technology in forensic science. *Journal of Forensic Sciences, 64*(3), 853–862. https://doi.org/10.1111/1556-4029.13956
Zhang, J. X., Yordanov, B., Gaunt, A., Wang, M. X., Dai, P., Chen, Y. J., Zhang, K., Fang, J. Z., Dalchau, N., Li, J., Phillips, A., & Zhang, D. Y. (2021). A deep learning model for predicting next-generation sequencing depth from DNA sequence. *Nature Communications, 12*(1).

Further reading

Arunkumar, R., Srinivasan, S., & Cazares, E. (2017). Application of artificial intelligence and machine learning algorithms to next-generation sequencing data. *bioRxiv*, 2017.06.11.153097.
Azolifesciences.com. (2022). *Human identification using next-generation DNA sequencing [Blog post]*. Retrieved March 1, 2023, from https://www.azolifesciences.com/article/Human-Identification-using-Next-Generation-DNA-Sequencing.aspx.
Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., … Bengio, Y. (2014). Generative adversarial nets. In *Advances in neural information processing systems* (pp. 2672–2680).
Illumina Inc. (n.d.). Forensic genomics | NGS solutions for forensic applications. Retrieved March 1, 2023, from https://www.illumina.com/areas-of-interest/forensic-genomics.html.
Kandel, S., Rubinstein, S., & Halperin, E. (2016). A novel machine learning tool for next generation sequencing variant identification. *PLoS One, 11*(9), e0162416. https://doi.org/10.1371/journal.pone.0162416
Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (pp. 1097–1105).
News Medical Life Sciences. (2022). *Sequencing could change research on sexual assault cases [News article]*. Retrieved March 1, 2023, from https://www.news-medical.net/news/20221216/Sequencing-could-change-research-on-sexual-assault-cases.aspx.
Sen, S. (2017). The applications and implications of next generation sequencing in precision medicine. *Annual Review of Genomics and Human Genetics, 18*.
Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need. In *Advances in neural information processing systems* (pp. 5998–6008). Curran Associates, Inc.

CHAPTER 28

Ethical issues of forensic application of NGS technology

Theresa M. DeAngelo[1] **and Kelly M. Elkins**[2]
[1]Maryland State Police Forensic Sciences Division, Pikesville, MD, United States; [2]Department of Chemistry, Forensic Science Program, Towson University, Towson, MD, United States

Introduction to forensic DNA analysis

Forensic identification using DNA typing has raised ethical concerns since it was introduced more than 35 years ago. The concerns include genetic privacy, consent, and using DNA to make predictions about a person's physical appearance (Elkins, 2013; Scwab et al., 2018; Wee & Mozur, 2021). Additional concerns center on racial disparities in databases and potentially developing leads to innocent people to crimes using familial DNA searches (Wickenheiser, 2019). While the earliest forensic DNA typing focused on comparisons of counts of autosomal DNA short tandem repeat (STR) loci repeats, modern forensic DNA typing methods analyze autosomal (a), X chromosome, and Y chromosome STR repeats via sequencing as well as single nucleotide polymorphism (SNP) genotypes and mitochondrial (mt) DNA haplotype variations (Jäger et al., 2017). After the initial genetic loci were chosen for analysis, they were sequenced, genotypes were documented, and population variations were recorded. At the time they were adopted, most STR loci were considered "junk" or noncoding DNA (Ohno, 1972).

Nowadays, it is known that many of the chosen STR and SNP forensic loci are in gene-coding regions and code for phenotype polymorphisms (Arneson et al., 2018; Bañuelos et al., 2021; Bañuelos et al., 2022; Pombar-Gomez et al., 2015; Shademan et al., 2021; Wyner et al., 2020). Indeed, many phenotype polymorphisms are explicitly typed in the newest forensic DNA typing kits. Forensic SNPs are classified into several types, including ancestry-informative SNPs (aiSNPs), phenotype-informative SNPs (piSNPs), and identity-informative SNPs (iiSNPs). Forensic DNA typing is used for not only identity typing but also for investigative leads and evidence analysis in criminal cases, including phenotype predictions, biogeographical ancestry predictions, monozygotic twin differentiation, and paternity and maternity lineage typing, missing persons, and cold cases (Elkins & Zeller, 2021; Weber-Lehmann et al., 2014) (Table 28.1). Additionally, whereas early forensic DNA typing was limited to a few loci, panels of hundreds or even thousands of loci are now easily analyzed with forensic DNA typing kits using next-generation sequencing (NGS). NGS data has been approved for upload to the Combined DNA Index System (CODIS) United States national DNA database since 2019

Next Generation Sequencing (NGS) Technology in DNA Analysis
ISBN 978-0-323-99144-5, https://doi.org/10.1016/B978-0-323-99144-5.00020-2
© 2024 Elsevier Inc.
All rights reserved.

Table 28.1 Types of forensic DNA typing and their applications.

Loci	Case
aSTRs	Identity, paternity, missing persons, cold cases
Y-STRs	Paternity, male lineage, missing persons, cold cases
X-STRs	Maternity, female lineage, missing persons, cold cases
mtDNA	Female lineage, missing persons, cold cases
iiSNPs	Identity, missing persons, cold cases
aiSNPs	Missing persons, cold cases, investigative leads
piSNP	Missing persons, cold cases, investigative leads
WGS or targeted NGS	Identity (including identical twins), missing persons, cold cases, historic cases, investigative leads

(Elkins & Zeller, 2021). Moreover, NGS DNA typing products are more specialized and yield more individualizing data, but they also require more analyst time and energy than ever before, as well as new data storage drives and mechanisms. Furthermore, whole genome sequencing (WGS) of an individual can be performed; WGS leads to the elucidation of the full genetic code for an organism or person. WGS can be invaluable when other DNA-typing approaches fail.

Instead of being limited to comparing STR repeats, forensic scientists employing NGS can compare sequence variation and facilitate genetic identification. The STR target loci have been determined to be more polymorphic than previously known through the identification of isoalleles, alleles of the same length but different sequences, many times differing by a single nucleotide base (Berglund et al., 2011; Gettings et al., 2017). NGS is facilitating the detection of mt heteroplasmy as well. Heteroplasmy and isoallelic variation can aid in DNA mixture deconvolution.

Ethics and forensic DNA typing

Since forensic DNA typing was introduced, bioethicists and citizens have raised numerous ethical and legal concerns (Elkins, 2013). Ethics can be defined as professional rules, norms, codes of conduct, and standards in a profession. The term law refers to a set of legal rules to regulate the actions of the members of a community. Ethics and laws should not be confused with morals, which are cultural norms and values governing behavior. Unintentional mistakes or instrument failures are not considered breaches of ethics. Herein, we will focus on ethical concerns of DNA typing.

Ethical concerns surrounding forensic DNA analysis range from including DNA profiles of individuals suspected of crimes in forensic databases after they are no longer suspects, keeping databases of missing persons from crime scene samples separate, obtaining consent for collecting and testing samples, using DNA typing only for crimes that are subject to forensic DNA typing, and expunging arrestee profiles uploaded to forensic databases if the persons are exonerated from the crime(s) (Elkins, 2013). Drawbacks of forensic STR DNA typing have been raised, including matching and arresting the wrong

twin in monozygotic twin pairs. Additional concerns, such as whether DNA phenotyping and ancestry predictions should be made, are more relevant than ever with the application of NGS. Several studies now point to forensic DNA kits uncovering disease phenotypes and the propensity for individuals to develop medical conditions later in life (Bañuelos et al., 2022). Other ethical issues raised with genotyping and sequencing applications include the reidentification of donors used for clinical or forensic population studies (Schwab et al., 2018; Wjst, 2010; Yang, 2019). Concerns with familial DNA analysis and the use of genetic genealogy have also been raised (Mateen et al., 2021). Ethical concerns raised include family members being used to lead law enforcement to their relatives to solve a crime without their consent and other privacy concerns (Court, 2018).

Understanding the differences between ethics, laws, and morals in determining the appropriate use and communication of DNA typing results requires education and practice that often extends beyond the university curriculum in chemistry and biology (Elkins & Fambegbe, 2020). Role-playing activities may help prospective law enforcement officers and forensic lab staff experience and consider different viewpoints for a case. For example, to engage with the content psychologically and understand concerns, a law enforcement officer or scientist may play the role of an officer going to a home or workplace to collect a DNA sample under a warrant in a genetic genealogy case, and another may play the role of the relative.

How does forensic genetic genealogy differ from familial DNA searches?

STR DNA typing data analyzed using government forensic databases to associate crime scene DNA with biologically related individuals using STR data is referred to as a familial DNA search. Software exists that searches databases for individuals that share alleles with the suspect contributor profile and identifies biologically related parents, siblings, and children. X- and Y-STRs are used to narrow the list to individuals that share paternal lineage. Familial DNA searching can identify an unknown offender and is regulated to be used in certain circumstances, but not as a routine step for criminal investigations. The names of close biological relatives are provided to investigators for further investigation. Concerns have been raised that the searches use relatives to identify the persons of interest in a criminal case (Katsanis, 2020). While this technique is banned in Maryland (our state) as of this writing, it is allowed in other states, including California, Colorado, New York, and Florida. In Montana, partial familial DNA searches are banned unless probable cause is found and a search warrant is issued.

Most recently, NGS DNA profiles are being used for genetic genealogy, the application of DNA testing to trace/locate individuals through their genetic family members using public, commercial genetic genealogy databases in cases for which National DNA Index System (NDIS) searches did not identify viable suspects (Wickenheiser, 2019). In investigative genetic genealogy (IGG), law enforcement officials upload forensic DNA SNP data from ForenSeq Kintelligence, or a similar commercial product, to public

genealogy databases, including GEDmatch, and compare to profiles from Ancestry, MyHeritage, and 23andMe, in order to scan the profiles of individuals who have opted in for law enforcement searches. The search can generate leads to genetic ancestors, including distant cousins, across multiple generations. Genetic genealogists generate family trees to identify modern descendants. This technique is also used in unique cases, such as murders, to identify serial killers and rapists.

On September 24, 2019, the US Department of Justice announced its interim policy on forensic genetic genealogical DNA analysis and searching (FGGS) with an effective date of November 1, 2019. Support statements were issued by the American Society of Crime Lab Directors (ASCLD), the Federal Bureau of Investigation (FBI), and the Consortium of Forensic Science Organizations.

On May 30, 2021, Maryland legislators passed a law regarding the use of IGG balancing public safety and the personal right to privacy that went into effect on October 1, 2021. Montana has passed similar restrictions. A search warrant is required to perform the search after the court issues a finding of probable cause. In Montana, public consumer databases cannot be searched unless the individual waives their right to privacy.

Maryland restricted the use of IGG to serious violent crimes, including committed or attempted murder, rape, a felony sexual offense, or a criminal act affecting public safety or national security (Ram et al., 2021; Taylor, 2021). In Maryland, the forensic sample must be biological material that investigators believe was deposited by a perpetrator and was collected from a crime scene, a person, an item, or a location connected to the criminal event, or the unidentified human remains of a suspected homicide victim. Traditional DNA typing, used to generate an STR profile from the forensic evidence and used in a CODIS search, must have failed to identify a known individual. Only public data from a direct-to-consumer (DTC) or public database that explicitly allows law enforcement to use it to investigate crimes should be used for the IGG search (Skeva et al., 2020).

Current concerns include the use of FGGS and IGG for lead generation and traditional investigation methods to follow up on or confirm a lead, especially when the FGGS or IGG is concealed in case notes, skirting legal and public concerns of its use.

Misuse of tools indicated for forensic use?

Off-label uses of products are common. For example, doctors can prescribe medicines and medical devices for off-label use if they have approved medical use for another condition. This has led to lives being saved. Parts and products can be used for purposes beyond the originally intended purpose. Similarly, IGG uses databases built by genealogists for law enforcement purposes.

In forensic DNA, products have been carefully tested and labeled with quality control seals and lot numbers. These labels are a means to convey that the products are reliable, accurate, and guaranteed for their intended forensic use. However, certified or accredited forensic labs are not the only purchasers or users of these products. Academic and

government-standard labs also buy forensic DNA products. Because forensic science is the application of science to law, the independent evaluation and validation of products by labs, both academic and forensic, and subsequent publication is important for admission to the court of law. However, off-label use is reportedly causing concern among companies. As of this writing, unlike for controlled substances, there are no restrictions or permissions needed to purchase forensic DNA products, including DNA extraction, quantitation, and DNA typing kits labeled for forensic use.

One application of off-label use is to develop population databases using various forensic kits to generate allele frequency databases for statistical calculations of genotypes. Standards labs have sequenced and analyzed forensic kit amplicons to independently check the products for their reliability in generating data and in producing accurate allele calls. STR kits targeting the same loci have been compared and assessed to ensure that they generate the same genotypes for a set of research-type forensic samples. Labs have validated the kits in their environment and workflow and determined threshold settings for stutter and limit of detection. From these studies, primer binding site mutations, nonconcordant genotypes at loci, loci with a high stutter, sequence variations, and additional STR alleles have been reported and published (Elkins & Zeller, 2021).

A second off-label use is the reanalysis of NGS data collected in support of law enforcement or medical diagnosis activities, such as the early detection of cancer. Reanalysis of the data by insurance companies could lead to the discovery of other genetic diseases or attributes and be used to deny patients insurance coverage in the future.

Separate from these research purposes, state labs in China have been accused of off-label use of forensic DNA typing kits to profile populations or ethnic groups (FTM, 2021). The groups have been reported to use DNA phenotype data from NGS to map human faces and profile or surveil the Uighur minority population (Wee & Mozur, 2021). Companies are considering additional purchasing restrictions on their products to avoid future misuse.

Personal genomics companies and consumer data in the forensic setting

Personal genomic companies offer DTC SNP panel DNA testing. The panels have been shown to be associated with various diseases and disease risk factors as well as personal characteristics. As these are uncovered, they are shared with consumers. Companies such as 23andMe, MyHeritage, Family Tree DNA, and AncestryDNA offer DNA typing of large SNP panels for ancestry predictions and genetic genealogy purposes. Other companies developed for the Russian, Chinese, Brazilian, and East Asian markets, such as GenoTek, Yoogene, Genera, and WeGene, respectively, predict talent, health, sport, and fitness traits. These companies have normalized genetic testing and seek to normalize the use of DNA for predictive health, but not forensic, purposes. However, they lack the rigorous, transparent, independent, and validated testing of their products and analysis of

their data that the public expects from forensic laboratories. Additionally, their owners have other planned uses for the data. For example, 23andMe has received large sums from the pharmaceutical company GlaxoSmithKline (GSK) to use the data in drug development, and GSK bought a 300 million USD stake in the company in 2018. Another example is how Blackstone, the $63 billion private equity giant, acquired Acenstry.com in 2020, which included individual consumer DNA samples as well as the databases that Ancestry.com has built for customers to track their family tree.

Consumers place trust in their data and findings and are using their data, such as that for BRCA gene mutations, to make life changing medical decisions, such as a preemptive mastectomy, when disease risk factors are identified. However, in some cases, this trust has been unfounded. Data has been found to be unreliable, wrong, or incorrectly interpreted. The tests are not designed to give accurate results for rare mutations. In the beginning, datasets were small as well, although they continue to grow.

Personal genomics companies also analyze SNPs from genes or gene regions that are associated with physical traits or personality characteristics. Some of these, including polymorphism predictors for hair color, eye color, skin tone, face shape, BMI, age, and height, are useful for forensic purposes (Elkins et al., 2023). The sequences need to be accurate in order to generate an accurate prediction. The predictions from commercial products developed for forensic applications are much more conservative. For example, forensic ancestry predictions using the ForenSeq kit are continental rather than regional by country, state, city, or town (Jäger et al., 2017). Facial prediction remains challenging, but web-based software such as MetaHuman generates images based on inputted characteristics (Elkins et al., 2023). Since full profiles are not always obtained, forensic genetic genealogists have begun imputing the sequences at SNPs based on the genotypes at nearby SNPs (microhaps). Their justification is that the prediction is used only for lead generation and will be confirmed with additional methods.

Forensic NGS and genetic genealogy

The application of DTC NGS data to genealogy to help solve crimes using IGG (Wickenheiser, 2019) is solving more cold cases faster than ever before. NGS was used to analyze a minor contributor in a sexual assault case (de Knijff, 2020) and identify Joseph James DeAngelo (no relation to the author) as the Golden State Killer in 2018 (Selk, 2018). Since then, hundreds of cold cases have been solved using IGG. IGG can be used to not only identify perpetrators but also to rule out suspects.

Privacy considerations are important for the application of DTC tools to cases (Bononi et al., 2020). Initially, GEDmatch required users to opt out of law enforcement searches. However, in 2020, they changed their policies to offer users the opportunity to opt into these searches rather than make them the default setting. Even with the changes, many people agreed to their data being used for investigative searches because of an innate desire

to "do good." Forensic genetic genealogists are committed to adhering to their guidelines and terms of service, following state and federal laws, and adhering to SWGDAM's IGG recommendations. The forensic raw DNA data for the suspect is used to search the database. GEDmatch does not reveal the raw data for chromosomal region matches in the database (Guerrini et al., 2021). After the database generates the matches, forensic genetic genealogists compare the results to nonDNA data, including social media and geographical location. Still, some consumers did not fully understand what they were agreeing to when opting into these searches. While the use of the GEDmatch database for investigative searches does not require contact with most individuals, when law enforcement needs to make contact, they obtain a warrant to get information, the names can be published in the news, and the person(s) may be asked to testify in court. Likewise, if a sample is requested from a suspect for probable cause, a court can decide if the request is reasonable. If a sample is retrieved from the trash, no warrant is required.

The use of these database searches for forensic investigative leads has caused concern for the general public that their DNA may be used to connect them with past or future major crimes and has led to rape victims refusing to engage with law enforcement (Wired, 2022). Many argue that there should be limitations on law enforcement's use of genetic genealogy databases to solve crimes (Guerrini et al., 2018). Most DTC companies do not delineate the process that enforcement agencies should use to request the use of their services for law enforcement purposes (Skeva et al., 2020). Those who are innocent and not related to a crime could become the center of an active investigation based on a match from a database search. This has caused the public's distrust of not just the companies providing DTC NGS and genetic genealogy but law enforcement and forensic investigators as well.

There are ethical concerns about NGS and genetic genealogy being used for surveillance and ideas about the development of a population database inclusive of everyone in the population. Additionally, there is concern regarding which cases law enforcement should be pursuing to keep the population safe. Genetic genealogy has been used to solve old cases for the identification of abandoned fetal human remains (Medina, 2022). While the circumstances of the cases—for example, if the fetus took a breath—are not clear from the skeletal remains alone, women are being arrested, and their cases are being brought to court to decide if they buried a miscarried fetus or killed or abandoned a baby years or decades earlier.

Next-generation sequencing and environmental impact

NGS, especially WGS, produces gigantic datasets. Data handling and storage are two of the significant problems challenging forensic labs seeking to implement NGS into their workflow (Alonso et al., 2017). Storing and backing up the data requires large servers or data centers. Manipulation and analysis of the data require computational cores, web-

based software, or computational tools. Servers and computer cores are energetically expensive. Although some are entirely powered by wind and solar farms, others are powered by fossil fuels. Moreover, powering servers uses energy resources that could be allocated for other uses. Google data centers use more power than entire countries (Forbes, 2021). Application of the new NGS technology to all criminal cases would present a huge national increase in energy usage.

Next-generation sequencing education needs

In 2008, Maryland became the only state in the US to ban the statewide DNA database for familial DNA searches. Thirteen years later, Maryland law HB240, "forensic genetic genealogical DNA analysis, searching, regulation, and oversight," authorizes forensic genetic genealogy analysis. The FGG law went into effect on October 1, 2021, and includes a requirement for training and certification for genealogists. On September 24, 2019, the US Department of Justice (DOJ) announced an interim policy on FGG DNA analysis and searching, effective November 1, 2019.

Although training and certification were regarded as important for inclusion, Maryland did not specify the training requirements to meet the certification requirements or the type of training and certification to meet the requirement. Numerous FGG courses of different lengths are available, including online and in-person options (Table 28.2).

Table 28.2 Forensic genetic genealogy courses.

Institution	Course/program title	Website
Brigham Young	Certificate in family history research	https://www.byupathway.org/certificate/family-history-research
Bureau of Justice Assistance, US DOJ	Forensic genetic genealogy virtual training	https://bja.ojp.gov/news/registration-open-forensic-genetic-genealogy-training
Boston University	Online certificate in genealogical research	https://genealogyonline.bu.edu/
Global Forensic and Justice Center at Florida International University	The fundamentals of investigative genetic genealogy	https://shop.gfjc.fiu.edu/product/the-fundamentals-of-forensic-genetic-genealogy/
National Criminal Justice Training Center at Fox Valley Technical College	Enhancing investigations through genetic genealogy	https://ncjtc.fvtc.edu/trainings/TR00011181/enhanced-investigations-genetic-genealogy
National Institute for Genealogical Studies	Forensic genealogy	https://www.genealogicalstudies.com/eng/courses.asp?courseID=36

Table 28.2 Forensic genetic genealogy courses.—cont'd

Institution	Course/program title	Website
Family Tree University	Genetic genealogy 101 (online)	https://university.familytreemagazine.com/courses/genetic-genealogy-101
University of New Haven	Forensic genetic genealogy online graduate certificate	https://www.newhaven.edu/lee-college/graduate-programs/certificates/forensic-genetic-genealogy/index.php
Genealogical Research Institute of Pittsburgh	Practical genetic genealogy (online)	https://www.gripitt.org/courses/practical-genetic-genealogy/
University of Strathclyde	Online postgraduate qualification in genealogical studies	https://www.strath.ac.uk/studywithus/centreforlifelonglearning/genealogy/geneticgenealogycoursesresearch/
Towson University	Honors seminar advanced topics: Forensic genealogy and historical research	https://catalog.towson.edu/undergraduate/course-descriptions/honr/

The University of New Haven's 12-credit online graduate certificate in forensic genetic genealogy, launched in January 2021, is arguably the most in-depth program available at the time of this writing. To earn the graduate certificate, students take semester-long courses in fundamentals of forensic biological evidence, forensic genetic genealogy, genealogical principles and methods, and FGG practicum, in which they complete a case research project. Other FGG courses and certificates are offered by Brigham Young University, the Bureau of Justice Assistance (BJA), Boston University (BU), the National Criminal Justice Training Center (NCJTC) at Fox Valley Technical College (FVTC), the Global Forensic and Justice Center (GFJC) at Florida International University (FIU), Family Tree University, the Genealogical Research Institute of Pittsburgh (GRIP), the National Institute for Genealogical Studies, the University of Strathclyde, and Towson University (TU). The courses offered by the BJA, the GFJC, Family Tree University, and GRIP are week-long courses, while the NCJTC is a 2-day training course. Most of the courses are online, although the NCJTC, TU Honors College, and GFJC courses are offered in person. The FVTC course is a 2-day course, while the GFJC and TU courses are 40-hr week-long or semester-long courses, respectively. Students and professionals can earn college credit from the university courses from BYU Global, BU, the University of New Haven, the University of Strathclyde, and TU.

Some specifics need to be clarified. Is credit-bearing education required to meet the certification requirements? Should there be a minimum grade point average? Should there be a certification exam to attain and maintain the certification to become an

investigative genetic genealogist? Maryland's law stipulates that a licensing program for laboratories performing SNP or another sequencing-based DNA testing on FGG evidence is established by October 1, 2022, and that a licensing program for individuals performing genetic genealogy is established by October 1, 2024.

Conclusion

As DNA analysis remains well-characterized and NGS technology has proven to be sound and acceptable for forensic applications, including upload to CODIS, the applications and uses of the products and technology need to be transparent and ethical. There is general consensus surrounding the use of IGG to solve murders, but the priority of other types of cases needs to be clarified. The data analysis and statistical tools' technological underpinnings and software calculations need to be open, transparent, and published. The communication of investigative results obtained from IGG needs to be clear and concise for all parties involved. This is not only legally sound but also ethical, so that the public maintains trust in law enforcement and forensic labs.

References

Alonso, A., Muller, P., Roewer, L., Willuweit, S., Budowle, B., & Parson, W. (2017). European survey on forensic applications of massively parallel sequencing. *Forensic Science International: Genetics, 29*, e23–e25. https://doi.org/10.1016/j.fsigen.2017.04.017

Arneson, D., Zhang, Y., Yang, X., & Narayanan, M. (2018). Shared mechanisms among neurodegenerative diseases: From genetic factors to gene networks. *Journal of Genetics, 97*, 795–806. https://doi.org/10.1007/s12041-018-0963-3

Bañuelos, M. M., Zavaleta, J. A., Roldan, A., Reyes, R. J., Guardado, M., Chavez Rojas, B., Nyein, T., Rodriguez Vega, A., Santos, M., Huerta Sanchez, E., & Rohlfs, R. (2021). Associations between forensic loci and neighboring gene expression levels may compromise medical privacy. *medRxiv*. https://doi.org/10.1101/2021.07.20.21260897

Bañuelos, M. M., Zavaleta, J. A., Roldan, A., Reyes, R. J., Guardado, M., Rojas, B. C., Nyein, T., Vega, A. R., Santos, M., Huerta-Sanchez, E., & Rohlfs, R. V. (2022). Associations between forensic loci and expression levels of neighboring may compromise medical privacy. *Proceedings of the National Academy of Sciences, 119*, 1–10. https://doi.org/10.1073/pnas.2121024119

Berglund, E. C., Kiialainen, A., & Syvänen, A.-C. (2011). Next-generation sequencing technologies and applications for human genetic history and forensics. *Investigative Genetics, 2*, 23. https://doi.org/10.1186/2041-2223-2-23

Bonomi, L., Huang, Y., & Ohno-Machado, L. (2020). Privacy challenges and research opportunities for genomic data sharing. *Nature Genetics, 52*, 646–654. https://doi.org/10.1038/s41588-020-0651-0

Court, D. S. (2018). Forensic genealogy: Some serious concerns. *Forensic Science International: Genetics, 36*, 203–204. https://doi.org/10.1016/j.fsigen.2018.07.011

de Knijff, P. (2020). Case study: How next generation sequencing resolved a difficult case, leading to the first criminal conviction of its kind. *Verogen*, 1–4. Retrieved November 27, 2020, from https://cdn2.hubspot.net/hubfs/6058606/Verogen-First-NGS-Court-Case-Study_Final_VD2019024_8.5x11-web.pdf?__hstc=238609695.bed74b81cf4041e42adad16833ab8584.1576870704888.1576870704888.1576870704888.1&__hssc=238609695.1.1576870704888.

Elkins, K. M. (2013). *Forensic DNA typing: A laboratory manual*. Waltham, MA: Elsevier.

Elkins, K. M., & Fambegbe, I. (2020). Case studies and methods for teaching professional ethics for forensic science students. *Journal of Forensic Science Education*, 2. https://jfse-ojs-tamu.tdl.org/jfse/index.php/jfse/article/view/17.

Elkins, K. M., Garloff, A. T., & Zeller, C. B. (2023). Additional predictions for forensic DNA phenotyping of externally visible characteristics using the ForenSeq and Imagen kits. *Journal of Forensic Sciences, 68*(2), 608–613. https://doi.org/10.1111/1556-4029.15215

Elkins, K. M., & Zeller, C. B. (2021). *Next generation sequencing in forensic science: A primer*. Boca Raton, FL: CRC Press.

Forbes (2021). Retrieved from www.forbes.com/sites/robertbryce/2020/10/21/googles-dominance-is-fueled-by-zambia-size-amounts-of-electricity/?sh=4d6a791768c9.

FTM. (2021). *Dutch company sells China DNA kits for ethnic cleansing*. Retrieved from www.ftm.eu/articles/qiagen-sells-dna-kits-china-ethnic-cleansing.

Gettings, K. B., Borsuk, L. A., Ballard, D., Bodner, M., Budowle, B., Devesse, L., King, J., Parson, W., Phillips, C., & Vallone, P. M. (2017). STRSeq: A catalog of sequence diversity at human identification short tandem repeat loci. *Forensic Science International: Genetics, 31*, 111–117. https://doi.org/10.1016/j.fsigen.2017.08.017

Guerrini, C. J., Robinson, J. O., Petersen, D., & McGuire, A. L. (2018). Should police have access to genetic genealogy databases? Capturing the Golden state killer and other criminals using a controversial new forensic technique. *PLoS Biology, 16*, e2006906. https://doi.org/10.1371/journal.pbio.2006906

Guerrini, C. J., Wickenheiser, R. A., Bettinger, B., McGuire, A. M., & Fullerton, S. M. (2021). Four misconceptions about investigative genetic genealogy. *Journal of Law and the Biosciences, 8*, 1–18. https://doi.org/10.1093/jlb/lsab001

Jäger, A. C., Alvarez, M. L., Davis, C. P., Guzmán, E., Han, Y., Way, L., Walichiewicz, P., Silva, D., Pham, N., Caves, G., Bruand, J., Schlesinger, F., Pond, S. J. K., Varlaro, J., Stephens, K. M., & Holt, C. L. (2017). Developmental validation of the MiSeq FGx forensic genomics system for targeted next generation sequencing in forensic DNA casework and database laboratories. *Forensic Science International: Genetics, 28*, 52–70. https://doi.org/10.1016/j.fsigen.2017.01.011

Katsanis, S. H. (2020). Pedigrees and perpetrators: Uses of DNA and genealogy in forensic investigations. *Annual Review of Genomics and Human Genetics, 21*, 535–564. https://doi.org/10.1146/annurev-genom-111819-084213

Mateen, R. M., Sabar, M. F., Hussain, S., Parveen, R., & Hussain, M. (2021). Familial DNA analysis and criminal investigation: Usage, downsides and privacy concerns. *Forensic Science International, 318*, 110576. https://doi.org/10.1016/j.forsciint.2020.110576

Medina, E. (2022). *Ohio woman convicted of murder in 1993 death of newborn*. Retrieved from www.nytimes.com/2022/04/05/us/geauga-child-verdict-ohio.html

Ohno, S. (1972). So much "junk" DNA in our genome. *Brookhaven Symposia in Biology, 23*, 366–370.

Pombar-Gomez, M., Lopez-Lopez, E., Martin-Guerrero, I., Carles, A. G. O., & De Pancorbo, M. M. (2015). Potential relationship between single nucleotide polymorphisms used in forensic genetics and diseases or other traits in European population. *International Journal of Legal Medicine, 129*, 435–443. https://doi.org/10.1007/s00414-015-1165-7

Ram, N., Murphy, E. E., & Suter, S. M. (2021). Regulating forensic genetic genealogy. *Science, 373*, 1444–1446. https://doi.org/10.1126/science.abj5724

Schwab, A. P., Luu, H. S., Wang, J., & Park, J. Y. (2018). Genomic privacy. *Clinical Chemistry, 64*, 1696–1703. https://doi.org/10.1373/clinchem.2018.289512

Selk, A. (2018). *The ingenious and 'dystopian' DNA technique police used to hunt the 'Golden State Killer' suspect*. Retrieved from www.washingtonpost.com/news/true-crime/wp/2018/04/27/golden-state-killer-dna-website-gedmatch-was-used-to-identify-joseph-deangelo-as-suspect-police-say/

Shademan, B., Biray Avci, C., Nikanfar, M., & Nourazarian, A. (2021). Application of next-generation sequencing in neurodegenerative diseases: Opportunities and challenges. *NeuroMolecular Medicine, 23*, 225–235. https://doi.org/10.1007/s12017-020-08601-7

Skeva, S., Larmuseau, M. H., & Shabani, M. (2020). Review of policies of companies and databases regarding access to customers' genealogy data for law enforcement purposes. *Personalized Medicine, 17*, 141–153. https://doi.org/10.2217/pme-2019-0100

Taylor, M. (2021). *Maryland, Montana become first states to restrict genetic genealogy searches*. Retrieved from www.forensicmag.com/577298-Maryland-Montana-Become-First-States-to-Restrict-Genetic-Genealogy-Searches/.

Weber-Lehmann, J., Schilling, E., Gradl, G., Richter, D. C., Wiehler, J., & Rolf, B. (2014). Finding the needle in the haystack: Differentiating 'identical' twins in paternity testing and forensics by ultra-deep next generation sequencing. *Forensic Science International Genetics, 9*, 42–46. https://doi.org/10.1016/j.fsigen.2013.10.015

Wee, S.-L., & Mozur, P. (2021). *China uses DNA to map faces, with help from the west*. Retrieved from www.nytimes.com/2019/12/03/business/china-dna-uighurs-xinjiang.html.

Wickenheiser, R. A. (2019). Forensic genealogy, bioethics and the Golden state killer case. *Forensic Science International: Synergy, 1*, 114–125. https://doi.org/10.1016/j.fsisyn.2019.07.003

Wired. (2022). *A rape survivor gave police her DNA. They linked her to another crime*. Retrieved from www.wired.com/story/a-rape-survivor-gave-police-her-dna-they-linked-her-to-another-crime/

Wjst, M. (2010). Caught you: Threats to confidentiality due to the public release of large-scale genetic data sets. *BMC Medical Ethics, 11*, 21. https://doi.org/10.1186/1472-6939-11-21

Wyner, N., Barash, M., & McNevin, D. (2020). Forensic autosomal short tandem repeats and their potential association with phenotype. *Frontiers in Genetics, 11*, 1–7. https://doi.org/10.3389/fgene.2020.00884

Yang, K.-M. (2019). Emerging ethical issues in managing real world data. *Healthcare Informatics Research, 25*, 237–238. https://doi.org/10.4258/hir.2019.25.4.237

CHAPTER 29

Ethical, legal, and social challenges of next-generation sequencing technologies (NGS) in forensic criminal identification

Helena Machado[1], Laura Neiva[1,2], Rafaela Granja[2] and Susana Silva[1,3]
[1]Institute for Social Sciences, University of Minho, Braga, Portugal; [2]Communication and Society Research Centre, University of Minho, Braga, Portugal; [3]Centre for Research in Anthropology, University of Minho, Braga, Portugal

Introduction

Next-Generation Sequencing (NGS), otherwise known as Massively Parallel Sequencing (MPS), or High Throughput Sequencing (HTS), promises to revolutionize the field of forensic science in supporting the activities of the criminal justice system. These techniques relate to DNA sequencing technologies through which multiple pieces of DNA are sequenced in parallel. NGS generates large amounts of data and thus it accelerates the broadening of DNA analysis beyond Short Tandem Repeats (STR) toward a wider array of genetic markers, in diverse applications such as forensic DNA phenotyping, ancestry assignment, and full mitochondrial analysis (Daniel et al., 2015; de Knijff, 2019; Hunter, 2018; Scudder et al., 2018b), while also opening new avenues for microbial forensics (Kuiper, 2016; Oliveira & Amorim, 2018).

Besides NGS's higher multiplexing capacity of combining different genetic markers, one other assured advantage of these techniques is to allow the search for "discernible uniqueness". They are therefore considered to be a real breakthrough in distinguishing identical twins (Amorim & Pinto, 2018). To summarize, NGS generates a vast amount of personal genomic data with unprecedented efficiency, speed, sensitivity, and depth of information that is expected to improve the fight against crime in the near future (Butler & Willis, 2020; de Knijff, 2019; Scudder et al., 2018a).

NGS amplifies the possibilities of genetic informativity by introducing direct or indirect analysis of coding regions of DNA, therefore generating very sensitive information, including inferences about health conditions. In this context, the boundary between "medical" and "forensic" has—or will soon become—a matter for dispute (Bonomi et al., 2020; Lefèvre, 2018; Machado & Silva, 2015; Wienroth et al., 2014; Williams & Wienroth, 2017).

Next Generation Sequencing (NGS) Technology in DNA Analysis
ISBN 978-0-323-99144-5, https://doi.org/10.1016/B978-0-323-99144-5.09001-6
© 2024 Elsevier Inc.
All rights reserved.

NGS has begun to be developed in bioengineering and biomedicine, but it has also been attracting the attention of forensic geneticists wanting to develop new methods for the identification of suspects of criminal acts (Alvarez-Cubero et al., 2017; Aly & Sabri, 2015; Ballard et al., 2020; Berglund et al., 2011; Budowle et al., 2017; Butler, 2015; Børsting & Morling, 2015; Imam et al., 2018; Kuiper, 2016; Oliveira & Amorim, 2018; Parson et al., 2016; Ryan et al., 2020). A recent European-wide online survey targeting forensic practitioners, involving 105 replies from participants representing police, governmental, academic, and private laboratories providing professional services in the field of forensic genetics in 32 European countries, showed that 73% already own an NGS platform or are planning to acquire one within the next 1−2 years (Gross et al., 2021).

Most of the existing literature about the ethical and social challenges of NGS focuses on the clinical medicine and health research domains (Ballard et al., 2020; Clarke, 2014; Curtis et al., 2019; Kochański & Demkow, 2015; Martinez-Martin & Magnus, 2019). However, it is recognized that more attention must be paid to the Ethical, Legal, and Social (ELSI) challenges of the uses of NGS in the field of criminal identification (Amorim & Pinto, 2018; Curtis et al., 2019; Samuel et al., 2018; Scudder et al., 2019). This chapter aims to contribute to filling this gap in ELSI literature by adopting the lens of a sociological understanding of NGS technologies as "promissory entities" which are framed by "expectations" toward the future development of forensic genetics (Tutton & Levitt, 2010). As suggested by Mike Fortun (2006), promissory narratives cannot be reduced to either hype, hope, or speculations, but occupy the uncertain, difficult space in between. The promise and related statements involving the future may be volatile, but they are also necessary to society and science, especially with regard to emergent technoscience in the forensic field (Wienroth, 2018b). This is particularly relevant given that "NGS poses massive challenges to science and society" (Bösl & Samida, 2021, p. 15).

Besides being aware of the potential of NGS technologies for supporting the activities of the criminal justice system, it is vital to attend to their ethical and social issues, including conflicts with human rights, especially in terms of privacy (Hong et al., 2015; Lippert et al., 2017; Martinez-Martin & Magnus, 2019; Ong et al., 2011; Schwab et al., 2018), and discrimination (Dupras et al., 2020; Duster, 2015; Moreau, 2019; Wauters & Van Hoyweghen, 2016). The security of genome-wide data and ethical and legal issues of data sharing should also be carefully and publicly debated (Bonomi et al., 2020). Such discussions are crucial to elaborate recommendations to support NGS's ethically responsible implementation and use in forensic practice, and within the criminal justice system more broadly, since the technology has advanced faster than the capacity to introduce safeguards.

Attempting to contribute to this debate, in the first sections of this chapter, we point to important challenges and potential issues related to the uses of NGS in the criminal justice system, by addressing the emerging ethical, legal, and social issues that deserve special attention and protection in legislation and forensic practices. In the final section, we detail the social and cultural aspects that surround the understanding of NGS's

technologies as "promissory entities" framed by "expectations". We argue that the perceived role of NGS for the future development of forensic genetics is at least partially explained by macro-level societal trends toward the "datafication" of society (Cukier & Mayer-Schoenberger, 2013; Sadowski, 2019; Van Dijck, 2014), expansion not only of massive surveillance in the digital world (Bellanova, 2017; Lyon, 2014), but also of surveillance capitalism (Lyon, 2017, 2019), and a trend toward an emphasis on the "actionable genome" (Houtman et al., 2021) that has led to a "participatory turn" in forensic genetics (Granja, 2020; Jasanoff, 2003). We conclude this chapter by calling for the need to develop guidelines for international bioethics focusing on NGS's uses and other genomic technologies in support of criminal justice activities in future years.

Gaps in legislation and regulation

Ethical and legal frameworks around forensic genetics are heavily influenced by the "non-coding" nature of the current markers, which allegedly has little connection to physical traits (except sex determination) (Curtis et al., 2019), and is perceived as not revealing sensitive information about individuals and their biological relatives. Much of the existing legislation on the collection and use of DNA in Europe, therefore, emphasize the restriction of forensic DNA to "non-coding" or "uninformative" regions of the human genome (Reed & Syndercombe-Court, 2016). However, this paradigm is being challenged by new research suggesting that STRs have a regulatory role in gene expression (Hunter, 2018).

In addition, forensics is in the process of a dramatic shift from these "non-coding" DNA methods. With the advent of high-throughput NGS, the generation of data at the nucleotide level beyond that of STR profiles alone has allowed laboratories to produce much more wide-ranging DNA information—for example, single nucleotide polymorphism (SNP), which can lead to greater discrimination between alleles (Gross et al., 2021; Slabbert & Heathfield, 2018). As more of the human genome is sequenced in a forensic context, then more attributes of the donor are revealed. These emerging opportunities for NGS technologies applications are problematic concerns, since it is unclear how well existing legal and regulatory mechanisms and encryption algorithms developed for genetic information are suited for the protection of large datasets in genome-wide analysis in forensics. Publicly available whole-genome datasets may become a resource for the reidentification of individuals or their relatives by linking their whole genome–derived STR profiles with common identifiers in DNA STR identification profile databases (Hong et al., 2015). This is problematic from a legal and ethical point of view since the growing use of NGS technologies might be considered a breach of genetic privacy by introducing direct or indirect analysis of coding regions of DNA. These require that sufficient safeguards must be implemented to guarantee, for example, the encryption of genetic data in order to protect the individuals' personal information (Skeva et al., 2020).

Considering that the donation of genetic profiles to genealogical databases and their subsequent use for criminal purposes entail consent issues, there are no legal provisions for destroying or returning the data when used for criminal investigations (Cho & Sankar, 2004). This new era of forensic genetics, therefore, exposes the inadequacy of current legislation (Curtis et al., 2019; Dupras et al., 2018; Samuel et al., 2018), and a "lack of an adequate legislative framework" (Butler & Willis, 2020, p. 354). As Curtis et al. (2019, p. 1484) stated "the law is already failing to catch up to advances in genetic technology." So, this emphasizes the need to think about current legislation to guarantee full and proper protection for genetic data (Curtis et al., 2019).

Access to databases that hold full-genome data

Genome-wide data, either in commercial databases aimed for recreational purposes or in databases for research and clinical purposes, continue to grow at a high rate, raising technical, ethical, and legal challenges (Borry et al., 2010; Chow-White et al., 2018; Kalokairinou et al., 2018; Kennett, 2019; Tutton, 2004). This means that DNA databases are becoming larger, in terms of the types of data included, and the number of people whose data are stored in them (Samuel et al., 2018).

The difficulties of the application of genome-wide data in forensics are of a statistical, computational, economic, and technical validation nature. Amorim and Pinto (2018, p. 104) emphasize the following question: "what to do with a substantial quantity of sensitive information that is not essential for identification purposes, but relevant for inferring traits that might be medically important or not?" The authors also claim that traditional ethical perspectives, as well as specific regulations, are clearly against the inference of information beyond identification purposes, and most current forensic DNA databases explicitly exclude any markers that can reveal physical traits or which are situated at coding regions (Amorim & Pinto, 2018).

However, while state-regulated DNA databases tend to limit the collection and archiving of data to DNA noncoding data, recent events have shown the increasing willingness of police agencies to access commercial databases for recreational purposes where citizens voluntarily upload genetic information to know more about their health and/or ancestry. Although these commercial tests have been validated by usage by genealogists, they have not been validated for forensic use. Nevertheless, since 2018 commercial databases aimed at recreational purposes have been used to identify suspects and missing persons by law enforcement agencies (Kennett, 2019). Such databases are not circumscribed by restrictive legislation and hold full-genome data (Bonomi et al., 2020; Granja, 2020; Kennett, 2019; Samuel et al., 2018; Samuel & Kennett, 2020a, 2020b; Schwab et al., 2018). Consequently, it is becoming increasingly challenging to guarantee that only noncoding DNA can be used for forensic purposes (Samuel et al., 2018).

The traditional ELSI debate in the field of forensic genetics usually focuses on forensic DNA databases created for criminal investigation purposes (Hindmarsh & Prainsack, 2010), so it does not capture issues that arise when data are used for other purposes beyond those for which it was originally collected, such as commercial DNA databases aimed for recreational purposes subsequently used in law enforcement investigations. Nonetheless, an ethically informed debate is urgently needed, considering that "in high-profile criminal cases it will be difficult to resist the political push to make data in these databases open to using in criminal investigations" (Samuel et al., 2018, p. e21).

In addition, there are several issues related to (the lack of adequate) consent from genetic genealogy database users (Samuel & Kennett, 2020a; Skeva et al., 2020). Specifically, the individuals may share their data without understanding the potential implications for themselves as well as their current and future blood relatives (Bonomi et al., 2020). Many people in genetic genealogy databases have not given informed consent to law enforcement, since most companies do not clearly inform users that their information may be subject to forensic analysis (Berkman et al., 2018). This has consequences both for those who voluntarily gave data for genealogy purposes which are now being used for forensic purposes, and for relatives who may be identified (Kennett, 2019). It is therefore necessary to point out the importance of obtaining consent from database users to allow access by law enforcement officers (Samuel & Kennett, 2020a). To this end, new legal standards should be designed to guarantee that the individuals are adequately informed about the future purposes for which their data can be used (Bonomi et al., 2020). Until now, how specific the consent should remain is an unresolved issue (Berkman et al., 2018).

Significant public concern has already been reported about access to and use of genetic genealogy data held in commercial databases (Berkman et al., 2018), but the discussion has been limited, disjointed, and unfocused (Butler & Willis, 2020; Curtis et al., 2019; Phillips, 2018). Among the most debated topics are issues involved in conceiving publicly accessible data as holding intelligence value for criminal investigation purposes, since certain criminal cases may justify access to noncriminal databases (Samuel et al., 2018). As Berkman et al. (2018, p. 1) stated, "a person might reasonably be surprised if his or her genealogic data were used in a criminal investigation, because that use is far afield from the original purpose for which the information was given". Forensic genetic genealogy also goes to the heart of the broader issue of surveillance, and the extent to which the privacy of citizens can be sacrificed to enforce the law (Curtis et al., 2019; Scudder et al., 2020), by expanding "surveillance capitalism" (Lyon, 2019, p. 66) where data is increasingly collected and accessed.

Finally, another topic of concern relates to the potential of the blurring of forensic and nonforensic databases. As forensic genetics and medical genetics converge toward genome sequencing, issues surrounding genetic data become increasingly intertwined and cannot be viewed in isolation (Machado & Silva, 2015; Samuel et al., 2018; Williams

& Wienroth, 2017), although less attention has been given to the use of genetic data for nonmedical purposes (Cho & Sankar, 2004). The speed of these developments has surprised many and demands a policy response to protect trust in medical genetics and in forensic genetics (Curtis et al., 2019), as the same data can be used both for medical and forensic purposes (Cho & Sankar, 2004). This is possible since some jurisdictions have made provisions allowing the use of health information by law enforcement in circumstances deemed to be reasonable (Curtis et al., 2019). This requires a change in the current legal standards to guarantee that access to genetic data is transparent and given oversight by independent bodies (Samuel et al., 2018).

Data protection

The recent use of genomic data for forensic purposes has brought privacy and data protection concerns to the attention of the general public (Bonomi et al., 2020). Data protection is achieved by enforcing restrictions on destruction and (unauthorized) access and actions, using technical tools. Data privacy is focused on the use and governance of personal data, and thus focused on the development of and adherence to administrative policies (Carter, 2019; Curtis et al., 2019; Kennett, 2019; Martinez-Martin & Magnus, 2019).

Considering that the traditional privacy models designed, for example, for health data provide limited protection, the same issues arise when we apply these models to NGS technologies in forensic genetics. The expansion of genealogy databases, which contain millions of DNA profiles (Kennett, 2019), and the increasing possibilities of accessing public personal information associated with genomic data (for example, family names, geographic data, and social media) (Bonomi et al., 2020), emphasize the importance of understanding how, in the future, a higher level of protection of individuals can be provided in the context of these new scenarios.

In addition, the recent tendency to adopt cloud services to support the amount of computational capacity required to analyze these large datasets raises new cybersecurity issues and threats to privacy and data protection. Since the threats to cybersecurity and privacy are increasing, this can pose significant risks to data donors, as their information can be accessed by unauthorized hacking activities (Carter, 2019). Data protection is urgently needed as there are increasing moves to expand informativity, for example, by using a combination of genomic data and phenotypic traits gathered from public databases, photos, and social media (Kim et al., 2018; Lippert et al., 2017).

The use of commercial databases aimed for recreational purposes to identify criminals, together with the advancing science of trait prediction in forensic investigations (Granja et al., 2020; Granja & Machado, 2020; Lippert et al., 2017; Scudder et al., 2018b; Wienroth, 2018a, 2018b), and other full-genome wide analysis are crucial advances that not only mark the start of a new era in forensic genetics (Machado & Granja,

2020) but also raise additional urgent challenges to the application of NGS technologies. As commercial and other human genetic information databases, containing genome-wide data, continue to grow at their current high rate, the possibility of obtaining a sample from a given individual and then matching it to a supposedly anonymous sample in a different database is no longer theoretical (Carter, 2019). Martinez-Martin and Magnus (2019) stated that it is estimated that in a few years with NGS and the growth of databases, it will be possible to identify almost anyone in the U.S.A. from a DNA sample, even if they had not voluntarily placed their genetic information in the public domain. Moreover, this poses data protection questions regarding whether the database users know that their DNA information will be searched by law enforcement in this way (Samuel & Kennett, 2020a). It is mandatory to respect individuals' rights to be adequately informed about how their data are being used (Skeva et al., 2020).

Algorithms, privacy, and discrimination

Whole-genome sequencing datasets contain information on current and future diseases and are essential for the future discovery of genetic etiologies of disease (Martinez-Martin and Magnus, 2019; Wauters and Van Hoyweghen, 2016) and have also been used to access biological relationships, including ancestry and parentage (Brück and Frieman, 2021), as well as reopening debates about genetic ancestry (Burmeister, 2021; Kohler, 2021; Lang and Winkler, 2021). To protect the DNA privacy of individuals participating in whole-genome sequencing, there needs to be constant vigilance to identify and protect against the possible misuse of information (Bonomi et al., 2020; Hong et al., 2015). In this regard, the issue of maintaining the confidentiality of data is particularly important. As outlined by Schwab et al. (2018), with NGS, profiled individuals may be reidentified through raw genetic data, even when social identifiers (name, date of birth, etc.) are removed and/or by identification through family members (see also Hong et al., 2015). This implies that if an individual decides to do genetic testing, the genetic privacy of the whole family (both current and future members) is implicated in such an option (Moreau, 2019).

NGS technologies generate large datasets that require software and algorithms to produce forensic analysis (Amorim & Pinto, 2018; Martinez-Martin & Magnus, 2019), which raises the potential for bias in the data used to construct the algorithms and bias in the algorithms themselves. If the dataset does not accurately reflect the population to which it is applied, then bias will be transferred to the outcomes generated by the algorithm. As pointed out by Martinez-Martin and Magnus (2019), the problem of bias presents a particular concern in the applications of NGS because of the historical lack of racial and ethnic diversity in human genetic and genomic research. This bias, together with police prejudice toward particular ethnic and racial minorities, raises the potential for differential treatment or abusive profiling of individuals or groups based on their

actual or presumed genetic characteristics (Crutchfield et al., 2012; Hunter, 2018; M'charek et al., 2013).

Forensic DNA databases usually overrepresent certain groups and social categories, such as racial and ethnic minorities (Duster, 2003; Skinner, 2013), since these groups are the preferential target of the actions of the criminal justice system. Although nonforensic DNA databases, such as medical, commercial, and research databases, usually present the opposite trend, lacking diversity in their composition, the increasing genetic informativity brought by NGS might nevertheless lead to new forms of discrimination. In this regard, two examples stand out. One regards the potential offered by NGS to augment the inference of phenotypic traits. By improving the predictions of forensic DNA phenotyping, and thereby increasing the visibility of racial or ethnic differences, NGS might contribute to the reproduction of stigmatization and reinforce the criminalization of certain groups (Lippert et al., 2017; Scudder et al., 2018a). Another example shows the contribution of NGS to forensic genealogy. By expanding genetic informativity, NGS might contribute to placing certain families under the lens of criminal investigation (Hong et al., 2015).

As outlined by Wauters and Van Hoyweghen (2016, p. 276), genetic discrimination refers to "the differential treatment of asymptomatic individuals or their relatives on the basis of their real or assumed genetic characteristics" which, up until recently, has been discussed mainly in terms of genetic diseases, usually referring to practices adopted by insurance companies and/or employers (Dupras et al., 2018, 2020). However, more recently researchers have been outlining how "epigenetic discrimination" might also significantly affect criminal investigations. As the interest in epigenetics applied to the field of criminal investigation rises (Haddrill, 2021; Hasnain, 2019; Vidaki et al., 2013), new avenues for differential treatment and abusive profiling also expand (Dupras et al., 2018). For instance, associations between epigenetic variants and deviant and criminal behaviors can maximize discrimination toward certain groups (Dupras et al., 2020).

The uses of algorithms and Big Data techniques[1] for NGS analysis might also lead to "black box" problems, where results or findings that arise from an automated tool are perceived as inherently more objective or accurate (Christin, 2020; Lefèvre, 2018; Martinez-Martin & Magnus, 2019). The study of algorithmic "black boxes" reveals that the discourse focused on technological magic might serve as a cloud of smoke that obscures the role of algorithms in reproducing these social processes of discrimination and surveillance (Brayne, 2017; Christin, 2020). As previously discussed in other studies on the application of Big Data and algorithms in criminal investigation (see e.g., Browning & Arrigo, 2021; Minocher & Randall, 2020; Sanders & Sheptycki, 2017), they are

[1] Generally characterized by aggregate, processing, and analyzing automatically massive datasets to produce algorithms to enhance the effectiveness and efficiency of decision-making (Christin, 2016; Kitchin, 2014; Stevens et al., 2018).

promoted under a veil of objectivity and veracity, building social expectations around their ability to produce unquestionable results (Boyd & Crawford, 2012). These widespread beliefs in Big Data techniques can mobilize arguments around NGS analysis as factual and capable of generating intelligence in a way that was not previously possible. Consequently, this "faith in the technology" (Lyon, 2014, p. 6) to sequence DNA automatically and rapidly to identify criminal suspects may expand social imaginaries associated with forensic genetics (Lynch et al., 2008) and shape expectations around NGS as capable of producing irrefutable truths. These arguments are fallacious and have ethical, legal, social, and cultural consequences. Firstly, they assume that these processes are characterized by transparency and participatory engagement, guiding neutral and objective decisions (Smith et al., 2017). Secondly, they establish trust in society to believe in the "epistemic capabilities of algorithms" (Aradau & Blanke, 2015). The analysis of these "black box" problems clarifies that all automated processes have errors and may reproduce erroneous conclusions (Martinez-Martin & Magnus, 2019). Therefore, it is necessary to begin "demythologizing" (Amorim, 2012, p. 259) algorithms and Big Data for NGS analysis as automated techniques capable of producing objective, neutral, and irrefutable results to fight crime, since these social expectations may influence the new directions that NGS application in forensic criminal identification is assuming.

Another problematic issue is the lack of transparency of industry developers who are generally reluctant to share information about the software they are designing and selling that would help forensic experts to evaluate the specific reasoning behind the outcomes generated (Amorim & Pinto, 2018; Granja & Machado, 2020; Hunter, 2018; Wienroth, 2018a). Increased transparency of these techniques and their creation processes is needed to open these "black boxes" of algorithms and illuminate how these mechanisms arise and create criminal suspicion (Leese, 2014). Considering that the producers of the technologies rarely answer questions about how they are created (Leese, 2014; Martinez-Martin & Magnus, 2019), and the error rates of NGS platforms are higher when compared to other techniques (Ballard et al., 2020), the involvement of forensic experts and scientists in problematizing NGS analyses in their early stages is urgent. It could create opportunities to open the "black boxes" of techniques and algorithms, and reduce their errors and bias.

Societal trends framing NGS technologies

Current and potential future uses of NGS technologies cannot be decontextualized from the historical, social, and cultural environment. Molecular technologies and genetic databases in the forensic context, including criminal identification, are often framed in terms of their societal benefits. They can help solve crimes by identifying suspects, but the uses of these technologies are also entangled with histories of coercion, racial surveillance, and the risk of stigmatization (Duster, 2003, 2006; Hindmarsh & Prainsack, 2010; Ossorio & Duster, 2005; Skinner, 2013).

Previous studies reveal that experts and scientists are aware of these ethical challenges and pose questions about how to use NGS techniques in an ethically sound way. These

ethical concerns, and the awareness of the ethical issues associated with NGS technologies, reveal two crucial aspects of the professional culture of the forensic geneticists' community. On the one hand, an adherence to the principles of anticipatory governance (Wienroth, 2018b) in which scientists seek to agree on appropriate practices in contexts tainted by controversy and uncertainty, ensuring the legitimacy of scientific knowledge and practice while invoking other values such as notions of accountability and transparency. On the other hand, invoking ethical challenges of NGS technologies allows an active practice of demarcation in relation to other agents active in the field of NGS techniques—clarifying the scientists' specific practices and rationales, while distancing themselves from private and commercial companies as well as from law enforcement agencies (Granja & Machado, 2020).

The exposure of the ethical challenges of the uses of NGS technologies is embedded in the sociality of science and in the way scientific work is legitimated. The awareness of privacy issues, as well as of discriminatory practices potentially generated by the flawed uses of NGS techniques employed by forensic practitioners, reveals the management of controversies, both of which are seen as ways to lend legitimacy and objectivity to scientific work (Granja & Machado, 2020; Machado & Granja, 2018). This sociological approach to published literature on the main ethical, legal, and social issues related to NGS in the forensic field allows us to understand how ethics plays a role in the ways in which scientists relate to each other and negotiate responsibilities both within and outside their fields of expertise (Knorr-Cetina & Mulkay, 1983; Latour, 1987).

NGS technologies in the forensic field are mostly perceived as "promissory entities" since despite being in an exploratory stage and barely used in routine casework, most forensic geneticists share the expectation that these technologies would be completely established in the forensic field scenario in the next decade, marking a "revolution". Its promissory character that underscores the revolutionizing aspect of NGS relates to the expectation that these techniques will allow the generation of a vast amount of personal genomic data with unprecedented efficiency, speed, sensitivity, and depth of information. This underpins the eagerness of health systems, corporations, and forensic initiatives to follow these productivities, efficiency, and power trends that characterize "surveillance capitalism" (Lyon, 2019, p. 66). Considering this macro context, NGS should be put into perspective as a materialization of this "surveillance capitalism" that marks the 21st century and is characterized as a collective desire to explore many and varied data, where people are attracted to the idea of ceding their genetic profile, and their information is increasingly accumulated. This makes evident how the arguments around NGS are embedded in "surveillance capitalism", as a general trend to achieve productive aims in neoliberal societies, where various participants are drawn toward their development, given the seductive nature of their features. As we have shown, this comes with a wide range of implications, not only in terms of how routine work is approached, but also in ethical, legal, and social considerations.

The conception of NGS as a "promissory entity" must be considered as it holds the power to guide the future of the field of forensic genetics in supporting criminal justice activities. As stated by Mike Fortun (2006, p. 156) "The rhetoric of the promise is everywhere in genomics, and it's all too easy and all too tempting to dismiss or overlook the real paradoxes of promising, and either take such statements at face value or dismiss them as 'mere hype.'" As several studies in the sociology of expectations have shown, expectations hold "generative" and "performative" power, as they guide activities, provide structure and legitimation, attract interest and foster investment (Borup et al., 2006; Brown & Michael, 2003; Konrad, 2006). By posting a desired future for these "promissory entities", and thereby mobilizing resources at the macro, meso, and micro levels, expectations are able to make such futures real (Borup et al., 2006). Since these social expectations may influence the new directions that NGS application in forensic criminal identification is assuming, we point out the importance of "demythologizing" (Amorim, 2012, p. 259) this technique as capable of producing objective, neutral, and irrefutable results to fight crime by opening its "black boxes." Specifically, we draw attention to the risks of assuming that NGS results are considered objective because they are produced by automated techniques such as Big Data. As extensive literature (Browning & Arrigo, 2021; Minocher & Randall, 2020; Sanders & Sheptycki, 2017) has already presented, data contains errors, and it is necessary to increase the transparency of these techniques (from their creators to their mode of use) in order to deconstruct these "black boxes" and to reduce their ethical and social consequences.

As far as current usage is concerned, NGS technologies tend to be conceived as useful for generating intelligence, and not evidence. This is framed by a broader trend in forensic science, already inaugurated by other technologies, such as forensic DNA phenotyping and familial searches, which have been shifting the focus of forensic science from the construction of evidence to the generation of intelligence considered valuable to criminal investigations (Machado & Granja, 2020; Wienroth, 2018b). In addition, the expectations about the use of NGS are also dependent upon professionals' access to resources and the type of work being conducted, the focus of which could be more centered either on research science or on forensic applied science. Professionals involved in research and who have access to a wide range of resources, owing to research funding and/or collaboration with equipment providers, tend to be more optimistic about NGS's current and future applications. On the contrary, forensic geneticists who are more involved with routine casework tend to describe these techniques as a time-consuming and expensive process with arguably added benefits for routine casework. Expectations about NGS's promising nature and the expected new possibilities associated with it can, therefore, be understood in relation to different locations of forensic geneticists within a developing network of innovation relationships (Machado & Granja, 2021).

Finally, it is important to consider the notion of responsibility. Nowadays, there is the prevalence of what some authors call the "actionable genome" (Houtman et al., 2021) that is leading to a "participatory turn" in forensic genetics (Granja, 2020). The widespread use of NGS technologies, each time at a lower cost for laboratories (clinical and medical or operating in the field of forensic services), makes the expansion of possibilities and choice enactment regarding the use of the human genome more evident. Houtman et al. (2021) argue that in present times "The genome is becoming actionable, meaning we can make decisions based on, or directly affecting, our personal genomic information. This is a result of recent developments in sequencing, interpreting, and editing the human genome" (Houtman et al., 2021, p. 1). From the perspective of the authors, this new capacity to make decisions about our own genome in several domains—ranging from recreational purposes to biobanking-related aims—must imply a strong sense of individual responsibility since decisions can also impact others and have wide-ranging implications in the future. Conversely, the authors also argue that genomic professionals equally hold the responsibility of enabling the process of opinion formation and informed deliberation. Considering the issue of individual responsibility, social responsibility, and the experts' accountability in genomics, it can be argued that the deliberation about the "actionable genome" is often not facilitated because of the "diffusion of responsibility" and also because guidance in the decision-making process is often lacking. While uses of NGS technologies can be put into practice by different stakeholders, such as medical organizations, research consortia, commercial companies, and a variety of biobanks, not only are the boundaries between clinical, research, public health, and commercial applications blurring, but also not all stakeholders are acting in the public's best interest (Houtman et al., 2021).

The "actionable genome" therefore implies that individuals might, indeed, make decisions about the uses of their genomic data that will affect not only them, but also their families, communities, and society at large. Nowadays, such decisions go beyond their option to know more about their health or ancestry—which were the aims underlining the growth of the direct-to-consumer genetic testing market (Borry et al., 2010; Chow-White et al., 2018; Tutton, 2004). More recently, the "actionable genome" also entails the decision of whether to make their genetic data available to law enforcement searches. According to Granja (2020), this marks a "participatory turn" in forensic genetics. Contrary to what previously happened, when individuals involved in law enforcement searches had some type of involvement with the criminal justice system, nowadays individual citizens interested in personal genomics and who have already purchased a direct-to-consumer genetic test have the possibility of choosing to make their data available for law enforcement actions.

This participatory turn is, nonetheless, still anchored in longstanding structures of power and inequality. Suspects or those who have been convicted within the criminal justice system, often from underprivileged and racialized communities, have their profiles

included in forensic DNA databases as mandatory (Hindmarsh & Prainsack, 2010), while consumers of recreational DNA databases, generally from economically privileged groups, have the option of choosing to make their profiles available to law enforcement altruistically (Granja, 2020).

Conclusion

The findings conclude that understanding ethical, legal, and social challenges, and ongoing monitoring of the applications of NGS in the criminal justice system is required. Practices related to data collection, storage, and use must be acceptable to gain the trust of the public. We argue that the perceived role of NGS for the future development of forensic genetics is at least partially explained by macrolevel societal trends toward "datafication" of society (Cukier & Mayer-Schoenberger, 2013; Sadowski, 2019; Van Dijck, 2014), expansion not only of massive surveillance in the digital world (Bellanova, 2017; Lyon, 2014), but also of surveillance capitalism (Lyon, 2017, 2019), and a trend toward an emphasis on the "actionable genome" (Houtman et al., 2021) that has led to a "participatory turn" in forensic genetics (Granja, 2020; Jasanoff, 2003).

The analysis of ethical, legal, and social implications should be integrated into genetic research, with the participation of scientists who can anticipate and monitor the full range of possible applications of the research from the earliest stages. Problematizing these techniques must be complemented by understanding how they can be used in ways that minimize harmful societal impacts, such as discrimination, stigmatization, and weakening data protection and genetic privacy. These discussions must reach out to forensic experts and other professionals in the criminal justice system in order to broaden the understanding of the social implications of these recent advances in the field of forensic genetics. The design and implementation of research direct the ways in which results can be used, and data and technology, rather than ethical considerations or social needs, tend to drive the use of science in unintended ways. Here we argue that geneticists should anticipate the ethical, legal, and social issues associated with nonmedical applications of NGS technologies. It will be necessary to develop comprehensive and standardized regulations to address these challenges.

Considering that the producers of these technologies are rarely involved in debates about their routine use, their problematization of NGS techniques could help to open the "black boxes" of these techniques to foster the understanding of their challenges more deeply. Based on their knowledge, it might be possible to understand how automation is working in NGS analyses to reduce the ethical, social, and cultural impact. It would also be necessary to undertake empirical analyses of the use of these techniques, to explore further: (1) How they could potentially be used and applied; (2) What are the proper rules about information access, storage, and security; and (3) What value they will have in identifying criminal suspects. Finally, it will be essential to create full awareness of

how NGS and their uses are different from previous DNA technologies in order to expose their potential to erode rights and expand surveillance capitalism. The current accelerated social trends around the implementation of new technologies such as NGS can be slowed down if, as a result of ethical, legal, and social debates, there is a call for reflection on questions such as: how necessary is the technology? How will it improve the fight against crime?

The consideration of the ethical, legal, and social issues of genetics research and the development of technology are very frequently perceived to slow down science. However, time spent making explicit the ever-present ethical, legal, and social issues and incorporating them into study design is better conceptualized as an integral part of the research process than as "extra" time, while also preventing serious obstacles that can also hamper or even halt research: absence of consideration of ethical issues can breed a deep distrust of scientists that can only hurt future efforts to carry out or raise funding for future research (Cho & Sankar, 2004). Finally, we argue for the urgent need for setting guidelines for international bioethics particularly oriented to NGS's uses and other genomic technologies in the future.

Acknowledgment

This work has received funding from a Consolidator Grant by the European Research Council (ERC), attributed to Helena Machado, under the European Union's Horizon 2020 research and innovation program (Grant agreement No. 648608). Complementary funding was received from the following grants/research programs: national funds through FCT – Fundação para a Ciência e a Tecnologia, I.P., under a Ph.D. Research Studentship with the reference 2020.04764.BD, attributed to Laura Neiva, under the Program Scientific Employment Stimulus – Institutional Call with the reference CEECINST/00157/2018, attributed to Rafaela Granja, and under the project UIDB/00736/2020 (base funding) and UIDP/00736/2020 (programmatic funding).

References

Alvarez-Cubero, M. J., Saiz, M., Martínez-García, B., Sayalero, S. M., Entrala, C., Lorente, J. A., & Martinez-Gonzalez, L. J. (2017). Next generation sequencing: An application in forensic science? *Annals of Human Biology, 44*(7), 581–592. https://doi.org/10.1080/03014460.2017.1375155

Aly, S. M., & Sabri, D. M. (2015). Next generation sequencing (NGS): A golden tool in forensic toolkit. *Archiwum Medycyny Sądoweji Kryminologii/Archives of Forensic Medicine and Criminology, 65*(4), 260–271. https://doi.org/10.5114/amsik.2015.61029

Amorim, A. (2012). Opening the DNA black box: Demythologizing forensic genetics. *New Genetics & Society, 31*(3), 259–270. https://doi.org/10.1080/14636778.2012.687083

Amorim, A., & Pinto, N. (2018). Big Data in forensic genetics. *Forensic Science International: Genetics, 37*, 102–105. https://doi.org/10.1016/J.FSIGEN.2018.08.001

Aradau, C., & Blanke, T. (2015). The (Big) Data-security assemblage: Knowledge and critique. *Big Data & Society, 2*(2), 1–12. https://doi.org/10.1177/2053951715609066

Ballard, D., Winkler-Galicki, J., & Wesoły, J. (2020). Massive parallel sequencing in forensics: Advantages, issues, technicalities, and prospects. *International Journal of Legal Medicine, 134*, 1291–1303. https://doi.org/10.1007/s00414-020-02294-0

Bellanova, R. (2017). Digital, politics, and algorithms: Governing digital data through the lens of data protection. *European Journal of Social Theory, 20*(3), 329–347. https://doi.org/10.1177/1368431016679167

Berglund, E. C., Kiialainen, A., & Syvänen, A. C. (2011). Next-generation sequencing technologies and applications for human genetic history and forensics. *Investigative Genetics, 2*(1), 1–15. https://doi.org/10.1186/2041-2223-2-23

Berkman, B. E., Miller, W. K., & Grady, C. (2018). Is it ethical to use genealogy data to solve crimes? *Annals of Internal Medicine, 169*(5), 1–2. https://doi.org/10.7326/m18-1348

Bonomi, L., Huang, Y., & Ohno-Machado, L. (2020). Privacy challenges and research opportunities for genomic data sharing. *Nature Genetics, 52*(7), 646–654. https://doi.org/10.1038/s41588-020-0651-0

Borry, P., Cornel, M. C., & Howard, H. C. (2010). Where are you going, where have you been: A recent history of the direct-to-consumer genetic testing market. *Journal of Community Genetics, 1*(3), 101–106. https://doi.org/10.1007/s12687-010-0023-z

Børsting, C., & Morling, N. (2015). Next generation sequencing and its applications in forensic genetics. *Forensic Science International: Genetics, 18*, 78–89. https://doi.org/10.1016/j.fsigen.2015.02.002

Borup, M., Brown, N., Konrad, K., & Van Lente, H. (2006). The sociology of expectations in science and technology. *Technology Analysis & Strategic Management, 18*(3–4), 285–298. https://doi.org/10.1080/09537320600777002

Bösl, E., & Samida, S. (2021). New sequencing methods. *TATuP - Zeitschrift Für Technikfolgenabschätzung in Theorie Und Praxis, 30*(2), 11–17. https://doi.org/10.14512/tatup.30.2.11

Boyd, D., & Crawford, K. (2012). Critical questions for Big Data. Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society, 15*(5), 662–679. https://doi.org/10.1080/1369118X.2012.678878

Brayne, S. (2017). Big Data surveillance: The case of policing. *American Sociological Review, 82*(5), 977–1008. https://doi.org/10.1177/0003122417725865

Browning, M., & Arrigo, B. (2021). Stop and risk: Policing, data, and the digital age of discrimination. *American Journal of Criminal Justice, 46*(2), 298–316. https://doi.org/10.1007/s12103-020-09557-x

Brown, N., & Michael, M. (2003). A sociology of expectations: Retrospecting prospects and prospecting retrospects. *Technology Analysis & Strategic Management, 15*(1), 3–18. https://doi.org/10.1080/0953732032000046024

Brück, J., & Frieman, C. J. (2021). Making kin. *TATuP - Zeitschrift Für Technikfolgenabschätzung in Theorie Und Praxis, 30*(2), 47–52. https://doi.org/10.14512/tatup.30.2.47

Budowle, B., Schmedes, S. E., & Wendt, F. R. (2017). Increasing the reach of forensic genetics with massively parallel sequencing. *Forensic Science, Medicine and Pathology, 13*(3), 342–349. https://doi.org/10.1007/s12024-017-9882-5

Burmeister, S. (2021). Does the concept of genetic ancestry reinforce racism? *TATuP - Zeitschrift Für Technikfolgenabschätzung in Theorie Und Praxis, 30*(2), 41–46. https://doi.org/10.14512/tatup.30.2.41

Butler, J. M. (2015). The future of forensic DNA analysis. *Philosophical Transactions of the Royal Society B: Biological Sciences, 370*(1674), 1–10. https://doi.org/10.1098/rstb.2014.0252

Butler, J. M., & Willis, S. (2020). Interpol review of forensic biology and forensic DNA typing 2016-19. *Forensic Science International: Synergy, 2*, 352–367. https://doi.org/10.1016/j.fsisyn.2019.12.002

Carter, A. B. (2019). Considerations for genomic data privacy and security when working in the cloud. *Journal of Molecular Diagnostics, 21*(4), 542–552. https://doi.org/10.1016/j.jmoldx.2018.07.009

Cho, M. K., & Sankar, P. (2004). Forensic genetics and ethical, legal and social implications beyond the clinic. *Nature Genetics, 36*(11), S8–S12. https://doi.org/10.1038/ng159

Chow-White, P., Struve, S., Lusoli, A., Lesage, F., Saraf, N., & Oldring, A. (2018). 'Warren buffet is my cousin': Shaping public understanding of Big Data biotechnology, direct-to-consumer genomics, and 23andMe on. *Twitter Information, Communication & Society, 21*(3), 448–464. https://doi.org/10.1080/1369118X.2017.1285951

Christin, A. (2016). From daguerreotypes to algorithms: Machines, expertise, and three forms of objectivity. *ACM SIGCAS - Computers and Society, 46*(1), 27–32. https://doi.org/10.1145/2908216.2908220

Christin, A. (2020). The ethnographer and the algorithm: Beyond the black box. *Theory and Society, 49*(5), 897–918. https://doi.org/10.1007/s11186-020-09411-3

Clarke, A. J. (2014). Managing the ethical challenges of next-generation sequencing in genomic medicine. *British Medical Bulletin, 111*(1), 17–30. https://doi.org/10.1093/bmb/ldu017

Crutchfield, R. D., Skinner, M. L., Haggerty, K. P., McGlynn, A., & Catalano, R. F. (2012). Racial disparity in police contacts. *Race and Justice, 2*(3), 179–202. https://doi.org/10.1177/2153368712448063

Cukier, K., & Mayer-Schoenberger, V. (2013). The rise of Big Data: How it's changing the way we think about the world. *Foreign Affairs, 92*(3), 28–40.

Curtis, C., Hereward, J., Mangelsdorf, M., Hussey, K., & Devereux, J. (2019). Protecting trust in medical genetics in the new era of forensics. *Genetics in Medicine, 21*(7), 1483–1485. https://doi.org/10.1038/s41436-018-0396-7

Daniel, R., Santos, C., Phillips, C., Fondevila, M., Van Oorschot, R. A. H., Carracedo, A., … McNevin, D. (2015). A SNaPshot of next generation sequencing for forensic SNP analysis. *Forensic Science International: Genetics, 14*, 50–60. https://doi.org/10.1016/j.fsigen.2014.08.013

Dupras, C., Joly, Y., & Rial-Sebbag, E. (2020). Human rights in the postgenomic era: Challenges and opportunities arising with epigenetics. *Social Science Information, 59*(1), 12–34. https://doi.org/10.1177/0539018419900139

Dupras, C., Song, L., Saulnier, K. M., & Joly, Y. (2018). Epigenetic discrimination: Emerging applications of epigenetics pointing to the limitations of policies against genetic discrimination. *Frontiers in Genetics, 9*(202), 1–6. https://doi.org/10.3389/fgene.2018.00202

Duster, T. (2003). *Backdoor to eugenics*. United Kingdom: Routledge.

Duster, T. (2006). Explaining differential trust of DNA forensic technology: Grounded assessment or inexplicable paranoia? *Journal of Law Medicine & Ethics, 34*(2), 293–300. https://doi.org/10.1111/j.1748-720X.2006.00034.x

Duster, T. (2015). A post-genomic surprise. The molecular reinscription of race in science, law and medicine. *The British Journal of Aociology, 66*(1), 1–27. https://doi.org/10.1111/1468-4446.12118

Fortun, M. (2006). For an ethics of promising, or: A few kind words about James Watson. *New Genetics & Society, 24*(2), 157–174. https://doi.org/10.1080/14636770500184792

Granja, R. (2020). Long-range familial searches in recreational DNA databases: Expansion of affected populations, the participatory turn, and the co-production of biovalue. *New Genetics & Society, 40*(3), 331–352. https://doi.org/10.1080/14636778.2020.1853515

Granja, R., & Machado, H. (2020). Forensic DNA phenotyping and its politics of legitimation and contestation: Views of forensic geneticists in Europe. *Social Studies of Science, 1–27*. https://doi.org/10.1177/0306312720945033

Granja, R., Machado, H., & Queirós, F. (2020). The (de) materialization of criminal bodies in forensic DNA phenotyping. *Body & Society, 27*(1), 60–84. https://doi.org/10.1177/1357034X20919168

Gross, T. E., Fleckhaus, J., & Schneider, P. M. (2021). Progress in the implementation of massively parallel sequencing for forensic genetics: Results of a European-wide survey among professional users. *International Journal of Legal Medicine, 135*(4), 1425–1432. https://doi.org/10.1007/s00414-021-02569-0

Haddrill, P. R. (2021). Developments in forensic DNA analysis. *Emerging Topics in Life Sciences, 5*(3), 381–393. https://doi.org/10.1042/ETLS20200304

Hasnain, S. E. (2019). Forensic epigenetic analysis: The path ahead. *Medical Principles and Practice, 28*(4), 301–308. https://doi.org/10.1159/000499496

Hindmarsh, R., & Prainsack, B. (2010). *Genetic suspects: Global governance of forensic DNA profiling and databasing*. Cambridge: Cambridge University Press.

Hong, C., Wang, J., Xing, C., Hwang, T. H., & Park, J. Y. (2015). Intersection of DNA privacy and whole-genome sequencing. *Clinical Chemistry, 61*(7), 900–902. https://doi.org/10.1373/clinchem.2014.235499

Houtman, D., Vijlbrief, B., & Riedijk, S. R. (2021). In the era of the actionable genome, everyone is responsible. *Trends in Biotechnology, 39*(7), 641–643. https://doi.org/10.1016/j.tibtech.2020.11.009

Hunter, P. (2018). Uncharted waters: Next-generation sequencing and machine learning software allow forensic science to expand into phenotype prediction from DNA samples. *EMBO Reports, 19*(3), e45810. https://doi.org/10.15252/embr.201845810

Imam, J., Shrivastava, P., Dixit, S., & Shrivastava, A. (2018). Future of DNA fingerprinting: Application of NGS in forensic science. In H. R. Dash, P. Shrivastava, B. K. Mohapatra, & S. Das (Eds.), *DNA fingerprinting: Advancements and future endeavors* (pp. 259–275). New York: Springer.

Jasanoff, S. (2003). Technologies of humility: Citizen participation on governing science. *Minerva, 41*, 223–244.

Kalokairinou, L., Howard, H. C., Slokenberga, S., Fisher, E., Flatscher-Thöni, M., Hartlev, M., … Borry, P. (2018). Legislation of direct-to-consumer genetic testing in Europe: A fragmented regulatory landscape. *Journal of Community Genetics, 9*(2), 117–132. https://doi.org/10.1007/s12687-017-0344-2

Kennett, D. (2019). Using genetic genealogy databases in missing persons cases and to develop suspect leads in violent crimes. *Forensic Science International, 301*, 107–117. https://doi.org/10.1016/j.forsciint.2019.05.016

Kim, J., Edge, M. D., Algee-Hewitt, B. F., Li, J. Z., & Rosenberg, N. A. (2018). Statistical detection of relatives typed with disjoint forensic and biomedical loci. *Cell, 175*(3), 848–858. https://doi.org/10.1016/j.cell.2018.09.008

Kitchin, R. (2014). Big Data, new epistemologies and paradigm shifts. *Big Data and Society, 1*(1), 1–12. https://doi.org/10.1177/2053951714528481

de Knijff, P. (2019). From next generation sequencing to now generation sequencing in forensics. *Forensic Science International: Genetics, 38*, 175–180. https://doi.org/10.1016/j.fsigen.2018.10.017

Knorr-Cetina, K. D., & Mulkay, M. (1983). *Science observed: Perspectives on the social study of science*. Beverly Hills, CA: Sage.

Kochański, A., & Demkow, U. (2015). Next generation sequencing—ethical and social issues. In U. Demkow, & R. Ploski (Eds.), *Clinical applications for next-generation sequencing* (pp. 301–307). Cambridge, Massachusetts: Academic Press.

Kohler, N. S. (2021). Negotiating Jewishness through genetic testing in the state of Israel. *TATuP - Zeitschrift Für Technikfolgenabschätzung in Theorie Und Praxis, 30*(2), 36–40. https://doi.org/10.14512/tatup.30.2.36

Konrad, K. (2006). The social dynamics of expectations: The interaction of collective and actor-specific expectations on electronic commerce and interactive television. *Technology Analysis & Strategic Management, 18*(3–4), 429–444. https://doi.org/10.1080/09537320600777192

Kuiper, I. (2016). Microbial forensics: Next-generation sequencing as catalyst: The use of new sequencing technologies to analyze whole microbial communities could become a powerful tool for forensic and criminal investigations. *EMBO Reports, 17*(8), 1085–1087. https://doi.org/10.15252/embr.201642794

Lang, A., & Winkler, F. (2021). Co-constructing ancestry through direct-to-consumer genetic testing. *TATuP - Zeitschrift Für Technikfolgenabschätzung in Theorie Und Praxis, 30*(2), 30–35. https://doi.org/10.14512/tatup.30.2.30

Latour, B. (1987). *Science in action: How to follow scientists and engineers through society*. Cambridge: Harvard University Press.

Leese, M. (2014). The new profiling: Algorithms, black boxes, and the failure of anti-discriminatory safeguards in the European Union. *Security Dialogue, 45*(5), 494–511. https://doi.org/10.1177/0967010614544204

Lefèvre, T. (2018). Big Data in forensic science and medicine. *Journal of Forensic and Legal Medicine, 57*, 1–6. https://doi.org/10.1016/j.jflm.2017.08.001

Lippert, C., Sabatini, R., Maher, M. C., Kang, E. Y., Lee, S., Arikan, O., … Venter, J. C. (2017). Identification of individuals by trait prediction using whole-genome sequencing data. *Proceedings of the National Academy of Sciences, 114*(38), 10166–10171. https://doi.org/10.1073/pnas.1711125114

Lynch, M., Cole, S. A., McNallly, R., & Jordan, K. (2008). *Truth machine: The contentious history of DNA fingerprinting*. Chicago: University of Chicago Press. https://doi.org/10.1017/CBO9781107415324.004

Lyon, D. (2014). Surveillance, Snowden, and Big Data: Capacities, consequences, critique. *Big Data and Society, 1*(2), 1–13. https://doi.org/10.1177/2053951714541861

Lyon, D. (2017). Digital citizenship and surveillance| surveillance culture: Engagement, exposure, and ethics in digital modernity. *International Journal of Communication, 11*, 824–842.

Lyon, D. (2019). Surveillance capitalism, surveillance culture and data politics. In D. Bigo, E. Isin, & E. Ruppert (Eds.), *Data politics* (pp. 64–77). United Kingdom: Routledge.

Machado, H., & Granja, R. (2018). Ethics in transnational forensic DNA data exchange in the EU: Constructing boundaries and managing controversies. *Science as Culture, 27*(2), 242–264. https://doi.org/10.1080/09505431.2018.1425385

Machado, H., & Granja, R. (2020). *Forensic genetics in the governance of crime*. Springer Singapore. https://doi.org/10.1007/978-981-15-2429-5

Machado, H., & Granja, R. (2021). *Genetic surveillance and crime control: Social, cultural and political perspectives*. United Kingdom: Routledge.

Machado, H., & Silva, S. (2015). Public participation in genetic databases: Crossing the boundaries between biobanks and forensic DNA databases through the principle of solidarity. *Journal of Medical Ethics, 41*(10), 820–824. https://doi.org/10.1136/medethics-2014-102126

Martinez-Martin, N., & Magnus, D. (2019). Privacy and ethical challenges in next-generation sequencing. *Expert Review of Precision Medicine and Drug Development, 4*(2), 95–104. https://doi.org/10.1080/23808993.2019.1599685

M'charek, A., Hagendijk, R., & Vries, W. D. (2013). Equal before the law: On the machinery of sameness in forensic DNA practice. *Science, Technology & Human Values, 38*(4), 542–565. https://doi.org/10.1177/0162243912453623

Minocher, X., & Randall, C. (2020). Predictable policing: New technology, old bias, and future resistance in Big Data surveillance. *Convergence, 26*(5–6), 1108–1124. https://doi.org/10.1177/1354856520933838

Moreau, Y. (2019). *Crack down on genomic surveillance*. Nature.

Oliveira, M., & Amorim, A. (2018). Microbial forensics: New breakthroughs and future prospects. *Applied Microbiology and Biotechnology, 102*(24), 10377–10391. https://doi.org/10.1007/s00253-018-9414-6

Ong, F. S., Grody, W. W., & Deignan, J. L. (2011). Privacy and data management in the era of massively parallel next-generation sequencing. *Expert Review of Molecular Diagnostics, 11*(5), 457–459. https://doi.org/10.1586/erm.11.34

Ossorio, P., & Duster, T. (2005). Race and genetics: Controversies in biomedical, behavioral, and forensic sciences. *American Psychologist, 60*(1), 115. https://doi.org/10.1037/0003-066X.60.1.11

Parson, W., Ballard, D., Budowle, B., Butler, J. M., Gettings, K. B., Gill, P., … Phillips, C. (2016). Massively parallel sequencing of forensic STRs: Considerations of the DNA commission of the international society for forensic genetics (ISFG) on minimal nomenclature requirements. *Forensic Science International: Genetics, 22*, 54–63. https://doi.org/10.1016/j.fsigen.2016.01.009

Phillips, C. (2018). The Golden State Killer investigation and the nascent field of forensic genealogy. *Forensic Science International: Genetics, 36*, 186–188. https://doi.org/10.1016/j.fsigen.2018.07.010

Reed, K., & Syndercombe-Court, D. (2016). *A comparative audit of legislative frameworks within the European Union for the collection, retention and use of forensic DNA profiles. Euroforgen-network of excellence*.

Ryan, L., Mathieson, M., Dwyer, T., Edwards, M., Harris, L., Krosch, M., … Taylor, E. (2020). Massively parallel sequencing as an investigative tool. *Australian Journal of Forensic Sciences*, 1–14. https://doi.org/10.1080/00450618.2020.1781251

Sadowski, J. (2019). When data is capital: Datafication, accumulation, and extraction. *Big Data & Society, 6*(1), 1–12. https://doi.org/10.1177/2053951718820549

Samuel, G., Howard, H. C., Cornel, M., van El, C., Hall, A., Forzano, F., & Prainsack, B. (2018). A response to the forensic genetics policy initiative's report "Establishing Best Practice for Forensic DNA Databases". *Forensic Science International: Genetics, 36*, e19–e21. https://doi.org/10.1016/j.fsigen.2018.07.002

Samuel, G., & Kennett, D. (2020a). Problematizing consent: Searching genetic genealogy databases for law enforcement purposes. *New Genetics & Society, 40*(3), 284–304. https://doi.org/10.1080/14636778.2020.1843149

Samuel, G., & Kennett, D. (2020b). The impact of investigative genetic genealogy: Perceptions of UK professional and public stakeholders. *Forensic Science International: Genetics, 48*(June). https://doi.org/10.1016/j.fsigen.2020.102366

Sanders, C. B., & Sheptycki, J. (2017). Policing, crime and 'Big Data'; towards a critique of the moral economy of stochastic governance. *Crime, Law and Social Change, 68*(1), 1–15. https://doi.org/10.1007/s10611-016-9678-7

Schwab, A., Luu, H., Wang, J., & Park, J. (2018). Genomic privacy. *Clinical Chemistry, 64*(12), 1696–1703. https://doi.org/10.1373/clinchem.2018.289512

Scudder, N., Daniel, R., Raymond, J., & Sears, A. (2020). Operationalising forensic genetic genealogy in an Australian context. *Forensic Science International, 316*(110543), 1–8. https://doi.org/10.1016/j.forsciint.2020.110543

Scudder, N., McNevin, D., Kelty, S. F., Funk, C., Walsh, S. J., & Robertson, J. (2019). Policy and regulatory implications of the new frontier of forensic genomics: Direct-to-consumer genetic data and genealogy records. *Current Issues in Criminal Justice, 31*(2), 194–216. https://doi.org/10.1080/10345329.2018.1560588

Scudder, N., McNevin, D., Kelty, S. F., Walsh, S. J., & Robertson, J. (2018a). Massively parallel sequencing and the emergence of forensic genomics: Defining the policy and legal issues for law enforcement. *Science & Justice, 58*(2), 153–158. https://doi.org/10.1016/j.scijus.2017.10.001

Scudder, N., McNevin, D., Kelty, S. F., Walsh, S. J., & Robertson, J. (2018b). Forensic DNA phenotyping: Developing a model privacy impact assessment. *Forensic Science International: Genetics, 34*, 222–230. https://doi.org/10.1016/j.fsigen.2018.03.005

Skeva, S., Larmuseau, M. H., & Shabani, M. (2020). Review of policies of companies and databases regarding access to customers' genealogy data for law enforcement purposes. *Personalized Medicine, 17*(2), 141–153. https://doi.org/10.2217/pme-2019-0100

Skinner, D. (2013). 'The NDNAD has no ability in itself to be discriminatory': Ethnicity and the governance of the UK national DNA database. *Sociology, 47*(5), 976–992. https://doi.org/10.1177/0038038513493539

Slabbert, N., & Heathfield, L. J. (2018). Ethical, legal and social implications of forensic molecular phenotyping in South Africa. *Developing World Bioethics, 18*(2), 171–181. https://doi.org/10.1111/dewb.12194

Smith, G. J., Bennett Moses, L., & Chan, J. (2017). The challenges of doing criminology in the Big Data era: Towards a digital and data-driven approach. *British Journal of Criminology, 57*(2), 259–274. https://doi.org/10.1093/bjc/azw096

Stevens, M., Wehrens, R., & de Bont, A. (2018). Conceptualizations of Big Data and their epistemological claims in healthcare: A discourse analysis. *Big Data & Society, 5*(2), 1–21. https://doi.org/10.1177/2053951718816727

Tutton, R. (2004). "They want to know where they came from": Population genetics, identity, and family genealogy. *New Genetics & Society, 23*(1), 105–120. https://doi.org/10.1080/1463677042000189606

Tutton, R., & Levitt, M. (2010). Health and wealth, law and order: Banking DNA against disease and crime. In R. Hindmarsh, & B. Prainsack (Eds.), *Genetic suspects global governance of forensic DNA profiling and databasing* (pp. 85–104). Cambridge: Cambridge University Press.

Van Dijck, J. (2014). Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology. *Surveillance and Society, 12*(2), 197–208. https://doi.org/10.24908/ss.v12i2.4776

Vidaki, A., Daniel, B., & Court, D. S. (2013). Forensic DNA methylation profiling—potential opportunities and challenges. *Forensic Science International: Genetics, 7*(5), 499–507. https://doi.org/10.1016/j.fsigen.2013.05.004

Wauters, A., & Van Hoyweghen, I. (2016). Global trends on fears and concerns of genetic discrimination: A systematic literature review. *Journal of Human Genetics, 61*(4), 275–282. https://doi.org/10.1038/jhg.2015.151

Wienroth, M. (2018a). Socio-technical disagreements as ethical fora: Parabon NanoLab's forensic DNA Snapshot service at the intersection of discourses around robust science, technology validation, and commerce. *BioSocieties, 15*(1), 28–45. https://doi.org/10.1057/s41292-018-0138-8

Wienroth, M. (2018b). Governing anticipatory technology practices. Forensic DNA phenotyping and the forensic genetics community in Europe. *New Genetics & Society, 37*(2), 137–152. https://doi.org/10.1080/14636778.2018.1469975

Wienroth, M., Morling, N., & Williams, R. (2014). Technological innovations in forensic genetics: Social, legal and ethical aspects. *Recent Advances in DNA & Gene Sequences (Formerly Recent Patents on DNA & Gene Sequences), 8*(2), 98–103.

Williams, R., & Wienroth, M. (2017). Social and ethical aspects of forensic genetics: A critical review. *Forensic Science Review, 29*(2), 145–169.

This page intentionally left blank

Index

Note: 'Page numbers followed by "*f*" indicate figures and "*t*" indicate tables'.

A

Acid phosphatase (AP) activity, 26
Actionable genome, 562
Adenosine triphosphate (ATP), 176
Aerosol filter tips, 61
Age-correlated hypermethylated CpG sites, 347
Age-prediction models, 350–351
  DNA methylation and, 320–323
Agouti signaling protein (ASIP), 297
Allele count ratio (ACR), 495–496
Alleles, 127, 195
  designation, 497–498
  frequency, 127
Alpha-amylase 1, 434
Alpha-naphthol, 427–428
Alpha-naphthyl phosphate, 427–428
Alternate light sources (ALS), 427–428
Amelogenin (Amel), 227
American Association of Blood Banks (AABB), 220
Amplification refractory mutation system polymerase chain reaction (ARMS-PCR), 186–187
AmpliSeq MH-74 plex research panel, 80
Analytical threshold (AT), 496–497
A-naphthyl phosphoric acid, 427–428
Ancestry informative markers (AIMs), 9–10, 44–45, 111, 115, 132–133, 320, 324, 480
Ancestry-informative SNPs (aiSNPs), 66, 134–139
Ancient DNA (aDNA) analysis, 196–197, 394, 515, 543–544. *See also* Forensic DNA analysis
  challenges, 198–199
  chemical modifications and degradation, 198–199
  contamination, 198
Ancient fecal samples, 197
Ancient mummies, mtDNA analysis of, 207–208
Ancient samples
  facility for, 200
  future perspective, 209–210
  mitochondrial DNA capturing and sequencing, 203–204
  next-generation sequencing (NGS)
    based on ancient sample mitochondrial DNA in evolutionary studies, 204–208
    data analysis, 204
    library preparation for, 203
    maternal genome analysis in, 199–204
  recent developments, 208–209
  sample preparation and DNA extraction from, 201–203
Animal DNA analysis, 47–48
Anthrax, 197
  letter attack, 92
Appearance traits (AT), 324
Applied Biosystems (ABI), 75, 291
  Precision ID mtDNA Whole Genome Panel, 487–488
Aqueous magnesium chloride, 427–428
Arabidopsis thaliana, 48
Artificial intelligence (AI), 533–534
  next-generation sequencing (NGS), 531, 535–536
    challenges in, 536
    for forensics, 531–532
Artificial neural networks (ANNs), 347–348
ASPA gene, 347
Atopobium vaginae, 436
Attenuated total reflectance Fourier transform infrared (ATR FTIR) spectroscopy, 279, 425–426
Autosomal DNA short tandem repeat, 539
Autosomal markers, 195
Autosomal SNP/InDel (Au-STR/InDel), 126–127
Autosomal STRs (Au-STR), 77, 126–127
  next-generation sequencing (NGS), 175–176, 495–500
    application, 176–178
    comparing loci amplified using different kits for, 178–181
    population studies, 181–183

B

Bacillus anthracis, 92, 376
Barcode of Life Data System (BOLD) database, 482
Base calling process, 86–87
Basic Tool (BT), 262–263
Bayesian inference (BI), 154–155
Bead-based normalization, 68–69
Beta-naphthol and Fast Garnet B, 427–428
Big Data techniques, 558–559, 561
Biogeographical ancestry (BGA), 49–50, 257–258, 320
Bioinformatics, 59, 104
Biological evidence, direct and indirect transfer of, 23–26, 33
Biological eyewitnesses, 292
Biological fluids, 30
Biological kinship, 219
Biological material, 22, 24
Biological sample, 311
Biological transfer, 24–25
Biomedical biobank, 521
Biomedical sciences, 534
Biospecimen and hacking threats, 519
Bioterrorism, 92
Black Blow Fly (*Phormia regina*), 482
"Black box" problems, 558–559
Bloodstains, 428–429
Bluemaxx BM500, 427–428
Body fluid identification (BFID), 358–359, 414–417
  by massive parallel DNA sequencing (MPS), 340–344, 341t, 415t
  techniques, 427
Body fluids, 418–419, 425
  detection/tissue differentiation, 436–437
Body height, 318–319
Bone powder, 61
Brentamine Fast Blue, 427–428
Bubonic plaque, 197
Bureau of Justice Assistance (BJA), 546–547
Bypassing codes of informed consent, 521–522

C

Caenorhabditis elegans, 357–358
Candidate molecular genetic markers
  allele frequency and genotype frequency, 127
  genetic markers located on sex chromosomes, 125–126
  haplotype frequency, 127–128
  Hardy-Weinberg equilibrium test, 124
  heterozygote (He) and polymorphism information content, 128
  linkage and linkage disequilibrium (LD), 124–125
  polymorphism parameters of genetic markers, 126–128
  population applicability of, 124–126
Capillary electrophoresis (CE), 35, 59, 85, 100, 123, 175, 221–222, 224, 245–253, 291, 337–338, 395, 398, 473–474, 485, 532
  based forensic DNA typing, 490
  based STR methods, 283–284
  data analysis, 6–7
  DNA sequencing platforms, 5–6
  forensic DNA analysis
    advantages of, 100
    limitations of, 100–101
  genomic analysis, 8
  kits for, 178–181
  metagenomics, 9
  methodology, 6
  next-generation sequencing (NGS) on, 11–15
    forensic genetics, 9–10
    genomics research, 7–8
  platform, 311
  recommendations, 15
  sequencing and imaging, 6
  short tandem repeats sequencing, 10–11
  targeted genomic resequencing, 8–9
  template preparation, 6
Catalytic tests, 428–429
Cell free fetal DNA (cffDNA), 106, 231–232
Cell-free DNA, 22–23
Chain termination method, 472
Chelex-100 technique, 62
Circulating cell-free fetal DNA (ccffDNA), 231
Circulating miRNAs, 357–358
Cleft lip, 266
Clinical exome sequencing, 476
Clostridium botulinum, 92
Coat hair of mammoth, mitochondrial DNA from, 205
Coding region single nucleotide polymorphisms (cSNPs), 244, 414–417

Cold cases, next-generation sequencing/ investigative genetic genealogy
methods, 452–459, 453t–458t, 459f–463f
results, 462–464
solved, 452–459
Combined parentage index (CPI), 221
Combined paternity index (CPI), 130
Commercial DNA databases, 555
Commercial NGS kits, for forensic analysis, 76–81
AmpliSeq MH-74 plex research panel, 80
ForenSeq DNA signature prep kit, 77
ForenSeq Kintelligence Kit, 77
ForenSeq MainstAY Kit, 77
ForenSeq mtDNA Control Region Kit, 78
ForenSeq mtDNA Whole Genome Kit, 78
illumina-verogen partnership, 76–78
ion ampliseq DNA phenotyping panel, 78
ion AmpliSeq HID Y-SNP research panel v1, 80
ion AmpliSeq pheno trivium panel, 79
ion AmpliSeq VISAGE-basic tool research panel, 80–81
precision ID ancestry panel, 79
precision ID globalfiler NGS panel v2, 80
precision ID identity panel, 79–80
precision ID mtDNA control region panel, 79
precision ID mtDNA whole genome panel, 79
Competitive grants, 505
Complementary DNA (cDNA), 6, 417–418
Complex traits, prediction of, 313
Compound biomarkers, 103
Compound genetic markers in microhaplotypes, types of, 103
Compound marker, 147–151
Contact trace DNA, 23
Contaminant DNA, 198
Continuous optimization, 392
Conventional SNP genotyping assays, 245
Conventional techniques, human body fluid detection, 387, 427–431. *See also* Molecular techniques
blood, 428–430
saliva, 430–431
semen, 427–428
sweat, 431
urine, 431
vaginal fluid, 430
Converge Software, 487–488
Copy number variation (CNV), 431–432
Criminal investigations, 103–104, 511
Criminal justice system, 551–552
Cumulative matching probability (CMP), 139–140
Cumulative probability of exclusion (CPE), 130
Cyclic reversible termination (CRT), 37–38
Cystic fibrosis transmembrane conductance regulating (CFTR), 435
Cytosine linked to guanine through phosphate group islands (CGIs), 320–321

D

Data analysis, 6–7
challenges and limitations, 109–110
Data interpretation, 267–268
Data protection, 556–557
Deep sequencing, 3
Degradation index (DI), 281–282, 517
Degraded DNA samples, 399
Degraded forensic samples, 517
Degraded samples, 398–400
Deletion/insertion polymorphisms present in flanking regions of STRs (DIP-STRs), 51
Deoxynucleotide triphosphates (dNTPs), 472
Dermcidin (DCD), 435
Developmental validation, 490–491
internal validation process, 495
process, 493–494
Dideoxy, 59
Di-deoxynucleotide triphosphates (ddNTPs), 3–4, 35, 472
Differentially methylated regions (DMRs), 337
Differentially methylated sites (DMSs), 337
Direct-to-consumer (DTC), 451–452, 514–515, 542
databases, 522
genomic services, 518–519
biospecimen and hacking threats, 519
false identification, 518–519
use of unproven or invalid testing methods, 519
Direct transfer of biological evidence, 23–26, 33
Disaster victim identification (DVI), 41, 104
Discrimination power (DP), 130, 222
DLX5 gene, 348
DNA, 395–398, 519
analysis methods, 185, 279
Scientific Working Group for DNA Analysis Methods validation guidelines for, 493–495
species identification by, 91–92

DNA (*Continued*)
barcoding method, 91, 481–482
based prediction, 312
damage and degradation, 392–393
databases, 487
DNA-DNA cross linking, 392–393
DNA-Protein cross-linking, 392–393
evidence, 394–395
application of next-generation sequencing (NGS) in, 398–403
extraction, 62–64, 377
from ancient samples, 201–203
forced collection of DNA samples from arrested individuals, 524–525
mixtures, 387–388, 401, 516–517
profile, 401–403
next-generation sequencing (NGS) techniques, 471
polymorphism, 357
profiles, 21–22, 27, 35, 301, 357–358
next-generation sequencing (NGS) kits and platforms for generating, 487–488
quantitation, 64–65
replication process, 392–393
testing, 22, 32
typing, 35, 387
advantages of microhaplotypes over, 105–106
techniques used in forensic DNA analysis, 100–101
DNA fingerprinting, 289, 359–362
DNA Identification Act (1994), 491
DNA markers, 222–224, 294
DNA methylation, 337, 347, 413–414
and age prediction, 320–323
analysis, 347
based assays, 337–338
high-resolution melting, 188–189
markers, 343
profiling, 351, 359
pyrosequencing, 189–190
DNA methyltransferase 1 (DNMT1), 321
DNA Primer Mix A (DPMA), 66, 502–503
DNA sequencing, 3–4, 99, 101–102, 376–377, 471
for forensic purposes, 85
platforms, 5–6
technology, 73, 551
DNA Solves, 515
DNA-STR Massive Sequencing & International Information Exchange (DNASeqEx), 401–402
DNase-free tubes, 61
Down syndrome, 266
Drowning, 367–368
DTC genomic industry, 518

E

EDAR gene, 319
Effective number of alleles (Ae), 105
Electrophoresis, 254–257
ELOVL2 gene, 324, 347, 349
Energy dispersive X-ray (EDX), 430–431
Epigenetics, 93–94, 388
age models, 347–348
age prediction model, 344
analysis, 45–46
approaches, 45–46
discrimination, 558
forensic applications
age estimation by massively parallel sequencing methods, 344–350
emerging massively parallel sequencing methods for forensics, 338–339
future considerations, 350–351
improvements in application of epigenetic markers, 350
marks, 320–321
methods, 351
modifications, 344
Epigenomics, 93–94, 337
Epithelial cells, 282
EpiTYPER, 348–349
Escherichia coli, 5–6, 368
Ethical, Legal, and Social issues (ELSI), 552
Ethical issues, 524–525
forced collection of DNA samples from arrested individuals, 524–525
lack of law protecting privacy, 525
Eumelanin, 264–265
European Network of Forensic Science Institutes (ENFSI), 485–486
European Standard Set (ESS), 181
Exome, 266
Exonic single nucleotide polymorphism, 244
External RNA Controls Consortium, 417–418
Externally visible characteristics (EVCs), 262–263, 293–299, 312, 480

eye color, 294–295
face feature, 298–299
forensic DNA phenotyping targets, 318–320
hair color, 295–297
skin color, 297–298
Eye color, 293–295
predictions, 326
Eyebrow color, 319
EyeColour 11 (EC11), 315

F

Face feature, 298–299
Facial morphology prediction, 298
Facial recognition software, 533–534
Facial structure, 293–294
False identification, 518–519
Familial DNA searches, forensic genetic genealogy differ from, 541–542
Familial searching, 479
FamilyTreeDNA, 515, 521, 543–544
Family Tree University, 546–547
Fast Red AL, 427–428
Fast-screening method, 281
Fatty acid-binding protein 5 (FABP5), 433
FDSTools, 7
Federal Bureau of Investigation (FBI), 491
quality assurance standards, 491–493
Federal funding, 505
FHL2 gene, 324
First generation sequencing, 35, 472
Fitzpatrick scale, 314
Fixation index (F_{ST}), 132
Flongle flow cells, 87–88
Florida International University (FIU), 546–547
Fluorescence resonance energy transfer (FRET), 40
Fluorescence spectroscopy, 64–65
ForenSeq DNA Signature Prep Kit, 77, 89, 177–178, 399–400, 475, 487, 502–503
ForenSeq Kintelligence Kit, 77
ForenSeq kits, 487
ForenSeq MainstAY Kit, 77, 487
ForenSeq mtDNA Control Region Kit, 78
ForenSeq mtDNA Whole Genome Kit, 78
thermo fisher scientific, 78–81
ForenSeq Prep Kit, 488
ForenSeq Universal Analysis Software, 487
Forensic analysis, 176–178, 426
commercial next-generation sequencing (NGS) kits used for, 76–81
Forensic anthropology purposes, forensic DNA phenotyping for, 325–326
Forensic applications
ethics and forensic DNA typing, 540–541
forensic DNA analysis, 539–540, 540t
forensic genetic genealogy differ from familial DNA searches, 541–542
misuse of tools, 542–543
of nanopore sequencing, 88–93
next-generation sequencing (NGS)
analysis of highly degraded and mixed samples, 258–260
education needs, 546–548
and environmental impact, 545–546
facial phenotyping, 264–266
genetic genealogy, 544–545
human ancestry and migration history, 262–263
individual identification, 261–262
low-copy DNA detection with NGS approach, 260–261
paternal lineage and kinship analysis, 266–267
of single nucleotide polymorphism markers, 258–267
personal genomics companies and consumer data in, 543–544
population genetics, 128–133
of population study, 122–123
Forensic biological examiner, 433
FORensic Capture Enrichment (FORCE), 260–261
Forensic community, 11–12
Forensic DNA, 542–543
advantages of capillary electrophoresis in, 100
analysis, 73, 99, 532
applications of microhaplotypes in, 106
DNA typing techniques used in, 100–101
ethics and forensic DNA typing, 540–541
future of, 48–49
limitations of capillary electrophoresis in, 100–101
profiling differs from forensic genetic genealogy, 511–512
Scientific Working Group for DNA Analysis Methods (SWGDAM) interpretation guidelines
for autosomal STR typing, 495–500

Forensic DNA (*Continued*)
for mitochondrial DNA analysis by forensic DNA testing laboratories, 500–502
testing, 294, 485
Forensic DNA databases, 41, 558
Forensic DNA laboratories
next-generation sequencing (NGS)
funding for forensic laboratories, 505–506
kits and platforms, 487–488
milestones for, 488–489
validation, 490–505
Forensic DNA phenotyping (FDP), 44, 257–258, 292, 312–326, 388, 478–479, 551, 561
DNA methylation and age prediction, 320–323
externally visible characteristics, 293–299, 318–320
final remarks, 326–327
for forensic anthropology purposes, 325–326
human pigmentation as starting point, 313–316
next-generation sequencing (NGS)
applicability of, 316–318
in forensics, 300–303
progression, 291–292
phenotypic-informative single nucleotide polymorphism in forensics, 292–293
prediction of complex traits, 313
short tandem repeats-based tool, 312–313
Visible Attributes Through Genomics initiative and future of, 324–325
Forensic entomology, applications of next-generation sequencing (NGS) technology in, 481–483
Forensic genetic genealogical DNA analysis and searching (FGGS), 542
Forensic genetic genealogy (FGG), 511
analysis, 77
differ from familial DNA searches, 541–542
forensic DNA profiling differs from, 511–512
law, 546
Forensic genetics, 122–123, 471–473
analysis, 243, 451
current limitations and problem solving in, 473–475
first generation sequencing, 472
landscape, 475
next-generation sequencing (NGS)
impact, 9–10
kits for, 11–12
timeline of important developments, 472f
second generation sequencing, 473
third generation sequencing, 473
Forensic identification, 539
Forensic investigations, 45
Forensic laboratories, 495–496
Forensic Laboratory for DNA Research (FLDO), 488
Forensic next-generation sequencing, 544–545
biological samples for
DNA extraction, 62–64
DNA quantitation, 64–65
library normalization, 68–69
library preparation, 65–68
next-generation sequencing (NGS), 59
sample preparation methods, 60–61
Sanger sequencing, 59
commercial systems, 48–50
Forensic parameters of population genetics, 123–133
Forensic phenotyping, 303
and ancestry studies, 43–45
Forensic samples using NGS technologies, panels developed for, 388–392
Forensic science, 122–123
hair analysis in, 279
next-generation sequencing (NGS) in, 59–60, 76–81
application, 388
Forensic scientists, 357, 359–362
Forensic setting, personal genomics companies and consumer data in, 543–544
Forensic single nucleotide polymorphism, 539–540
Forensic testing, prevalent applications of nucleic acid analysis in, 477f
Forensics, 316–317, 368–369, 531, 534, 553
current limitations and problem solving in forensic genetics, 473–475
forensic genetics, 471–473
limitations of single nucleotide polymorphism analysis in, 267–268
massively parallel sequencing methods for, 338–339
misuse of tools indicated for, 542–543
next-generation sequencing (NGS), 475–479, 531–532
forensic entomology, 481–483
in forensics and issue of privacy, 479–481

phenotypic-informative single nucleotide polymorphism in, 292–293
Fourier transform infrared (FTIR) spectroscopy, 279
Fox Valley Technical College (FVTC), 546–547
*FRAS*1 gene, 319
F test, 152–153
Full mitochondrial analysis, 551
Full-genome data, access to databases, 554–556
Full-genome wide analysis, 556–557

G

*GATA*3 gene, 319
GEDMatch, 519, 521
GEDmatch PRO, 515
GenapSys Sequencing Platform, 39
Gene frequencies, 121, 158–159
Genealogical connections, 480
Genealogical databases, 522–524, 524f
Genealogical Research Institute of Pittsburgh (GRIP), 546–547
Genealogy databases, 556
GeneReader system, 39
Genes on chromosome, 195
Genetic differentiation index (F_{ST}), 151–152
Genetic genealogy, 512f, 526–527, 544–545
  case studies, 525–526
    golden state serial killer or case of Joseph DeAngelo, 525–526
    Phoenix canal killings, 525
  challenges in using genetic genealogy, 515–526, 526t
    direct-to-customer genomic services, 518–519
    errors, 515–518
    ethical issues, 524–525
    intrusion to privacy, 519–524
  forensic DNA profiling differs from forensic genetic genealogy, 511–512
  in forensics, 526–527
  works, 512–515
    basic steps involved in investigative genetic genealogy, 514f
Genetic relationships
  AI-SNPs
    ancestral inference, 164
    phylogenetic analyses, 161–162
    principal component analysis analyses, 160–161
    structure analysis, 162–164
  investigation of, 159–164
  population differentiation measures of 17 continental populations, 159–160
Genetics, 294
  database, 462–464
  discrimination, 558
  genealogical databases, 515
  markers, 195, 243
    located on sex chromosomes, 125–126
    polymorphism parameters of, 126–128
  material, 61
  privacy, 553
  surveillance, 522
Genome Analyzer, 39
Genome Sequencer 20, 74–75
Genome-wide association studies (GWAS), 116, 264–265, 285, 290–291, 312
Genome-wide data, 554
Genome-wide DNA methylation profiling, 190–192
Genomic DNA (gDNA), 79
Genomics, 101, 106
  impact of next-generation sequencing (NGS) on genomics research, 7–8
  privacy, 519–521
GenoTek, 543–544
Genotype frequency, 127
Genotyping
  DNA, 221–222
  next-generation sequencing (NGS), 225–226
    kits for parentage and kinship testing, 226–227
  technology, 224–227
    capillary electrophoresis, 224
Geolocation in United States, application of microbial DNA analysis in, 382–383
Gestational week (GW), 231
GlaxoSmithKline (GSK), 543–544
Global Forensic and Justice Center (GFJC), 546–547
Glycosidic bonds, 199
GridION system, 75–76

H

Hair analysis
  in forensic science, 279
  next-generation sequencing (NGS) application for, 283–286
Hair color, 293–297, 314
  predictions, 326

Hair evidence samples, 279
Hair pigmentation, 295–296
Hair root, 281
Hair samples, 279
Hair shaft, 280
Haplotype analysis, 325
Haplotype frequency, 127–128
Haplotype reconstruction, 109
Hardy-Weinberg equilibrium test, 124
Healing process, 364
HERC2 gene, 313–314
Heteroplasmy, 90–91, 280
Heterozygosity, 128
Heterozygote (He), 128
High-resolution melt analysis targeting microRNA, 413–414
High Throughput Sequencing (HTS), 551
High-throughput sequencing methods, 5–6
Highly variable (HV), 280
HIrisPlex system, 44, 265–266, 296
HIrisPlex-S, 297–298, 317
  multiplex assays, 316
  system, 315
  tool, 265–266
HiSeq X10, 39
Histone deacetylases (HDACs), 321
Histone H3 lysine 9 methylation (H3K9me), 321
Hominin from South Siberia, mtDNA analysis of, 207
Homopolymers, 88–89
Housekeeping genes, 417–418
Human ancestry, 262–263
Human biological specimens, 392
Human body, 22–23, 426–427, 436
  existing technologies for human body fluid detection, 427–438
Human genome, 7–8
Human Genome Project, 291
Human identification (HID), 67–68
  methods, 381
  microhaplotypes as tool for, 102–106
Human pigmentation, 313–316
  phenotypes, 264–265
Human skeletons, 197
Humic acid, 377
Hybrid methods, 285–286
Hypervariable (HV) regions, 377, 487–488
  HV1, 90–91, 196
  HVII, 90–91, 144–145, 196, 502
  HVIII, 144–145
  HVR1, 144–145, 198, 502

I

Identical by descent (IBD), 221, 512–514
Identical by state (IBS), 221, 522
  probing, 523–524
  tiling, 522–523
Identity testing single nucleotide polymorphism, 261–262
Identity-informative single nucleotide polymorphism (iiSNPs), 66, 539–540
Illumina, 37–39, 531
  sequencing technology, 182–183, 381
Illumina HiSeq 2500, 75
Illumina Human Methylation 450 K bead array, 342
Illumina Infinium HumanMethylation450 bead chip, 344
Illumina MiSeq, 340–342, 344, 399, 475
  FGx Forensic Genomics System, 76–77, 338–339
  platform, 284–285
Illumina-verogen partnership, 76–78
Immunochromatographic strip tests, 425–426
Immunoglobulin G (IgG), 432, 451, 512–515
Indirect contact, 23–24
Indirect transfer of biological evidence, 23–26, 33
Individual identification SNPs (IISNPs), 11, 49
Induced errors, 515–516
Influenza, 197
Informed consent, bypassing codes of, 521–522
Insertion/deletion polymorphisms (InDels), 222
  markers, 223, 397
Institut National de Police Scientifique (INPS), 488
Intelligent BioSystems CRT platform, 39
Interim Policy on Forensic Genetic Genealogical DNA Analysis and Searching, 517
Internal organ tissues, 366–367
Internal Transcribed Spacer (ITS2), 482
Internal validation, 478–479, 490–491
  process, 495
International Genetics of Anthropometric Traits (GIANT), 318–319
International Society for Forensic Genetics (ISFG), 398, 486

Intrusion to privacy, 519–524
bypassing codes of informed consent, 521–522
via genealogical databases, 522–524
genetic surveillance, 522
genomic privacy, 519–521
Invalid testing methods, use of unproven or, 519
Investigative Genetic Genealogy (IGG), 451–459, 511
Investigator DIPlex kit, 397
Ion AmpliSeq DNA phenotyping panel, 78
Ion AmpliSeq HID Y-SNP research panel v1, 80
Ion AmpliSeq pheno trivium panel, 79
Ion AmpliSeq VISAGE-basic tool research panel, 80–81
Ion Chef System, 487–488
Ion detection, 176
Ion GeneStudio S5 System, 38, 253, 487–488
Ion PGM System, 11
Ion PGMTM System, 49
Ion S5 system, 339, 473, 475, 503–504
Ion sphere particles (ISPs), 67–68
Ion Torrent, 475
Genexus System, 38
MPS assay, 316–317
Personal Genome Machine, 203–204, 339, 473, 531
platforms, 475
systems, 75, 85
Ion TorrentTM, 473
IrisPlex
genetic test, 295
system, 44, 295, 314
iSeq 100 System, 75
Isoalleles, 224
ITGA2B gene, 347

K

Kinship analysis, 266–267
DNA markers used for parentage and, 222–224
principles of parentage and, 221
samples used for parentage and, 221–222
Kinship testing, 220–221, 234–235
commercially available next-generation sequencing (NGS) kits for, 226–227
through history, 219–220
KIT ligand (KITLG), 297
*KLF*14 gene, 324, 348

L

Lactobacillus
L. crispatus, 436
L. jensenii, 436
Law enforcement agencies, 487, 540–541
Library normalization process, 68–69
Life Technologies SOLiD, 531
Likelihood ratios (LR), 234–235, 419–420
Lineage-informative SNPs, 134–139
Linkage disequilibrium (LD), 106–107, 109, 124–125
Liquid chromatography-mass spectrometry (LC-MS), 431
Locard's exchange principle, 23
Long interspersed element 1 (LINE-1), 189–190
Low-copy DNA detection, 260–261
Low copy number DNA, 394
Low-quality forensic samples, 517
Low–template DNA, 387–388, 394
evidence, 398–400

M

Machine learning (ML), 533–535
algorithms, 86–87
next-generation sequencing (NGS), 531, 535–536
challenges in, 536
for forensics, 531–532
Magnetic beads, 62–63
Magnetite-modified silicon oxide magnetic beads, 62–63
Mammoth, mitochondrial DNA from coat hair of, 205
Maryland legislators, 542
Massive(ly) parallel sequencing (MPS), 3, 36, 81, 175, 225, 245–253, 312–313, 337–338, 387–388, 480, 551. *See also* Next-generation sequencing (NGS)
assays, 318
for forensics, 338–339
body fluid identification, 340–344
Illumina NovaSeq, 338
Illumina's MiSeq FGx sequencing system, 338–339
Ion Torrent PGM and Ion S5, 339
Oxford Nanopore Technologies MinION device, 339

Massive(ly) parallel sequencing (MPS) (*Continued*)
improvements in application of epigenetic markers via, 350
methods, 350–351
age estimation by, 344–350
technology, 106, 111
uniparental disomy in parentage testing, 230–231
Maternal DNA (mtDNA) markers, 195
Maternal genome analysis in ancient samples, next-generation sequencing (NGS) for, 199–204
Matrix metalloproteinase-11 (MMP-11), 432–433
Maxam–Gilbert sequencing, 291
Maximum likelihood (ML), 154–155
Maximum parsimony (MP), 154–155
MC1R gene, 313–314
Mean absolute deviation (MAD), 344
Melanocortin 1 receptor gene (MC1R), 295–296
Melanosomes, 297
Messenger RNA (mRNA), 3, 413–414
analysis, 93, 418–419
body fluids, 418–419
future of, 419–420
identification, 414–417
marker, 432–433
sequencing, 414, 417–418
Metabolically active cells, 199
Metagenomics, 9
classification, 376–377
MetaHuman (Web-based software), 544
Method validation, 490
Methyl (-CH_3), 320–321
Methyl-CpG-binding domain (MBD), 322
Methylation-sensitive restriction enzymes (MSREs), 322
5-Methylcytosine (5 mC), 321
Microbial analysis, 92–93
Microbial DNA
analysis
geolocation in United States, 382–383
next-generation sequencing (NGS) methods for, 381–382
extraction methods, 377–381
profiling, 383
quantitation methods for, 381
Microbial forensics, 46–47, 92
comparative account of techniques used for detection of body-fluids, 439t–443t
technological advancements in, 438
Microbial-based body fluid detection, 436–437
signature microflora found on different tissues/body fluids, 437t
Microbiome, 10
Microhaplotypes (MHs), 80, 101–102, 110–112, 145–147, 224, 300–301
analysis
advantages, 105–106
ancestry inference, 104
applications, 106
challenges and limitations of, 104–105
criminal investigations, 103–104
definition of, 102–103
disaster victim identification, 104
DNA sequencing, 101–102
DNA typing techniques used in forensic DNA analysis, 100–101
ethical considerations for NGS-based microhaplotypes analysis, 113–114
forensic DNA analysis, 99
future directions and opportunities in NGS-based microhaplotypes analysis, 110–113, 116–117
next-generation sequencing (NGS) for, 106–108
potential applications of NGS-based microhaplotypes analysis, 113
as tool for human identification, 102–106
types of compound genetic markers in, 103
validation of NGS-based microhaplotypes analysis, 108–110
future directions, 113
genetic marker, 124–125
markers, 224
Microhaps, 102, 300–301
MicroRNAs (miRNAs), 188, 357–359, 367–368, 425–426
markers, 365t, 434
organ tissue identification using, 366–367
profiling, 185
and time since death determination, 362–364
wound age estimation using, 364–366
Microsatellites, 243–244
Migration history, 262–263
Mini-sequencing (MiSeq), 75, 290–291, 348–349, 381
FGx Forensic Genomics System, 50, 66, 301–302, 392, 473

FGx sequencing instrument platform, 487
FGx Sequencing System, 50, 85
FGx system, 488
Sequencer, 399
Minimum postmortem interval (mPMI), 481
MinION device, 85
MinION flow cells, 87–88
MinION technique, 75
MiniSTRs, 395–396
MIR29B2C gene, 324
miRNA-706, 367–368
Mitochondrial DNA (mtDNA), 73, 90–91, 126–127, 144–145, 244, 302, 387, 476–478, 487
analysis, 35, 280–281, 500–501
of ancient mummies, 207–208
of Hominin from South Siberia, 207
capturing and sequencing, 203–204
from coat hair of mammoth, 205
genotyping, 90–91
markers, 195
of Paleo-Eskimo, 206
profiling, 357
Scientific Working Group for DNA Analysis Methods(SWGDAM), 500–502
sequencing, 47
typing, 396
Mitochondrial genome sequencing, 185–187
Molecular biology techniques, 198
Molecular filters, 202–203
Molecular techniques, 431–435. *See also* Conventional techniques
blood, 432–433
microbial-based body fluid detection, 436–437
rapid techniques, 435–437
saliva, 433–434
semen, 431–432
sweat, 435
technological advancements in microbial forensics, 438
urine, 434–435
vaginal fluid, 433
Monozygotic (MZ) twins, 185, 190–191
differentiation, 45
discrimination of, 359–362
using next-generation sequencing (NGS)
detecting DNA methylation with pyrosequencing, 189–190
DNA analysis methods, 185
genome-wide DNA methylation profiling, 190–192
high-resolution melting to determine DNA methylation, 188–189
mitochondrial genome sequencing, 185–187
nuclear genome sequencing, 187–188
RNA profiling, 188
Morgan's law, 122
Motor protein, 86–87
Mucin 4 (MUC4), 433
Mucin-5B, 434
Multidimensional scaling (MDS) analysis, 151, 153
Multinomial linear regression (MLR) algorithm, 314
Multiple alleles, 127
Multiple internet search engines, 452–459
My-Forensic-Loci-queries (MyFLq), 401
Mycobacterium tuberculosis, 197
MyHeritage, 515, 519, 543–544

N

NanoDrop, 381
Nanopore sequencing, 85–88, 93, 392
challenges, 94–95
forensic applications of, 88–93
mitochondrial DNA genotyping, 90–91
nonhuman forensics, 91–93
short tandem repeat genotyping, 88–89
single nucleotide polymorphism genotyping, 89–90
opportunities, 93–94
epigenetics and epigenomics, 93–94
messenger RNA analysis, 93
techniques, 473
Nanopores, 75, 87–88
systems, 40
National Criminal Justice Training Center (NCJTC), 546–547
National DNA Index System (NDIS), 487, 491, 520–521, 541–542
National Institute of Justice (NIJ), 505–506
Neanderthals, genetic history of, 206–207
Neighbor-joining (NJ) algorithms, 161
Neisseria subflava, 436
Neurology community, 476
Neutral theory of molecular evolution, 122

Next-generation sequencing (NGS), 36, 59, 110–112, 175–176, 185, 224–226, 280–281, 290–291, 337–338, 357–358, 387–388, 417–418, 426f–427f, 451, 471, 475–479, 485, 517, 539–540, 551
access to databases, 554–556
advantages for STR and SNP genotyping, 254–258
algorithms, privacy, and discrimination, 557–559
analysis of challenging samples
challenging DNA evidence, 394–395, 398–403
DNA damage and degradation, 392–393
panels developed for forensic samples, 388–392
previous types of technologies, 395–398
technology in forensic science, 388
artificial intelligence and machine learning help with, 535–536
autosomal STR analysis, 176–178
background, 425–427
based microhaplotypes analysis, 101–102
based on ancient sample mtDNA in evolutionary studies, 204–208
ancient mummies, 207–208
coat hair of mammoth, 205
genetic history of Neanderthals, 206–207
Hominin from South Siberia, 207
Paleo-Eskimo, 206
based sequencing platforms, 388
challenges of, 112, 536
data management and analysis, 112
low coverage depth and variability, 112
read errors and artifacts, 112
comparing loci amplified using different kits, 178–181
conventional genotyping assays next-generation sequencing (NGS) technology, 245–258
current status of, 12–15, 253–254
data analysis, 204, 382
data protection, 556–557
development of, 73–76
commercial next-generation sequencingkits used for forensic analysis, 76–81
second-generation next-generation sequencing systems, 74–75
third-generation sequencing systems, 75–76
use in forensic sciences, 76–81
discrimination of monozygotic twins, 359–362
DNA profiles, 541–542
education needs, 546–548
forensic genetic genealogy courses, 546t–547t
and environmental impact, 545–546
ethical considerations for, 113–114
existing technologies for human body fluid detection, 427–438
conventional techniques, 427–431
molecular techniques, 431–435
forensic applications, 40–48
animal and plant DNA analysis, 47–48
epigenetic analysis, 45–46
forensic phenotyping and ancestry studies, 43–45
microbial forensics, 46–47
mitochondrial DNA sequencing, 47
RNA analysis, 46
short tandem repeats and single nucleotide polymorphism analysis, 41–43
SNP markers using, 258–267
in forensic entomology, 481–483
forensic next-generation sequencing commercial systems, 48–50
advantages and challenges, 50–52
in forensics, 9–10, 300–303, 531–532
funding for, 505–506
and issue of privacy, 479–481
science, 59–60
future directions, 116–117
and opportunities in, 110–113
gaps in legislation and regulation, 553–554
guidelines for responsible use of, 114
hair analysis, 283–286
high-throughput capability, 8
impact on genomics research, 7–8
improvements to, 112–113
developing and optimizing computational tools, 112–113
enhancing coverage depth and consistency, 112
future directions in microhaplotype analysis, 113
reducing read errors and artifacts, 112
initial commercial next-generation sequencing kits for forensic genetics, 11–12
kits and platforms for generating DNA profiles, 487–488
kits for parentage and kinship testing, 226–227
library preparation for, 203
low-copy DNA detection with, 260–261

for maternal genome analysis in ancient samples, 199–204
methods, 196, 350–351
micro RNAs (miRNA)
and body fluid identification, 358–359
biomarkers, 360t–361t
and drowning, 367–368
and sepsis, 368
and time since death determination, 362–364
for microhaplotypes analysis, 106–108
milestones for next-generation sequencing (NGS) in forensic casework, 488–489
mitochondrial DNA (mtDNA)
ancient mummies analysis, 207–208
Hominin from South Siberia analysis, 207
coat hair of mammoth, 205
of Paleo-Eskimo, 206
obligations of scientists, researchers, and policymakers, 114
opportunities for future research in, 116–117
organ tissue identification using miRNA, 366–367
peer-reviewed publications on validation studies of, 502–505
platforms, 36–40, 123
commonly used markers of population study on, 133–151
learning practices of population study based on, 157–164
second-generation sequencing, 37–39
third-generation sequencing, 39–40
population studies of autosomal STRs using, 181–183
potential applications of, 113, 368–369
potential limitations, 369
progression to, 291–292
risks of, 113
sample preparation process for, 60–61
Sanger sequencing *vs.*, 59
societal trends framing, 559–563
state-of-the-art, 116
technologies, 3, 73, 101, 245–253, 312–313, 376, 471–472
traditional capillary electrophoresis-based assays *vs.*, 254
validation of, 108–110
challenges and limitations related to data analysis, 109–110
haplotype reconstruction, 109
quality control, 108–109
next-generation sequencing-based DNA analysis methods, 490–505
read alignment, 109
single nucleotide polymorphism (SNP) calling, 109
statistical inference, 109
workflow, 107–108
alignment, 108
data analysis, 108
library preparation, 107–108
sample preparation, 107
sequencing, 108
wound age estimation using miRNA, 364–366
NextSeq 1000–2000, 75
NextSeq 550, 75
NGSpeciesID, 91–92
NIST standard reference materials, 492–493
Non-coding DNA methods, 552
Nonallelic signals, 498
Nonforensic DNA databases, 558
Nonhuman forensics, 91–93
microbial analysis, 92–93
species identification by DNA analysis, 91–92
Noninvasive prenatal paternity testing (NIPPT), 222–223, 231–234
NonSanger-based technology, 253
Nonspecific DNA quantification methods, 381
Nuclear DNA (nuDNA), 280
analysis, 281–282
Nuclear genome sequencing, 187–188
Nucleic acids, 6, 245–253, 362–363
prevalent applications of nucleic acid analysis in forensic testing, 477f
Nucleotides, 5–6, 245–253
Nuffield Council study (2007), 521

O

OB mRT-PCR technique, 436
Obligate allele, 221
OCA2 gene, 313–314
Open-source software program, 7
Organ tissue identification using miRNA, 366–367
Osteopontin, 434–435
Othram Inc., 338
Oxford Ancestors, 515

Oxford Nanopore Technologies (ONT), 85, 339, 471
MinION device, 339

P

PacBio sequencing, 40
PacBio's single-molecule real-time sequencing, 392
Paleo-Eskimo, mitochondrial DNA of, 206
Panels developed for forensic samples using NGS technologies, 388–392, 389t–391t
Parentage testing, 130–131, 226–234
commercially available NGS kits for, 226–227
DNA markers used for parentage and kinship analysis, 222–224
genotyping technology, 224–227
massive parallel DNA sequencing (MPS), 230–231
noninvasive prenatal paternity testing, 231–234
parentage and kinship testing through history, 219–220
postnatal parentage testing, 227–230
principles of parentage and kinship analysis, 221
samples used for parentage and kinship analysis, 221–222
Partial least squares discriminant analysis (PLS-DA), 262–263
Paternal lineage, 266–267
Paternity, 219
test, 121
Paternity Index (PI), 105, 122–123
Paul Coverdell National Forensic Science Improvement Act Program, 505
PDE4C gene, 347
Peak height ratio (PHR), 495–496
Percentage-based threshold, 496–497
Personal Genome Machine (PGM), 6, 75, 144–145, 203–204
Personal genomics companies, 544
and consumer data in forensic setting, 543–544
Phenol-chloroform-isoamyl alcohol (PCIA), 62
Phenotype-informative single nucleotide polymorphism (PISNPs), 66, 134–139, 145–147, 539–540
in forensics, 292–293
Pheomelanin, 264–265
Phoenix canal killings, 525
Phylogenetic tree, 154–155
Physical evidence recovery kit (PERK), 30
Plant DNA analysis, 47–48
p-nitroaniline, 427–428
Point heteroplasmy, 185–186
Polilight, 427–429
Polymerase chain reaction (PCR), 35, 62–63, 99, 101–102, 186–187, 198, 220, 289–290, 357, 377
based short tandem repeat systems, 28
dependent methods, 243
Polymerase chain reaction-capillary electrophoresis (PCR-CE) methods, 3
Polymorphism information content (PIC), 126–128
Polymorphism parameters of genetic markers, 126–128
Population genetics, 122, 124–126
forensic application analysis based on, 128–133
individual identification, 129–130
inference of ancestry informative, 132–133
parentage testing, 130–131
forensic parameters of, 123–133
study, 121
Population Informative Multiplex for Americas (PIMA), 320
Population studies, 121–122
of autosomal STRs, 181–183
commonly used markers of population study, 133–151
forensic applications of, 122–123
forensic parameters of population genetics, 123–133
forensic population study, 121–123
genetic theories of, 122
learning practices, 157–164
assessing contribution
gene frequency statistics
investigation of genetic relationships, 159–164
statistical analytical methods, 151–157
Postmortem cells, 199
Postmortem interval (PMI), 362–363
Postmortem submersion interval (PMSI), 482–483
Postnatal parentage testing, 227–230
PostPCR high-resolution melt (PCR HRM), 188–189
Power of discrimination (DP), 129
PowerPlex Fusion System, 401–402
PowerSeq 46GY System, 487
PowerSeq CRM Nested System, 286, 487, 489, 504–505

PowerSeq product, 487
Prader-Willi syndrome, 266
Precision ID ancestry panel, 79
Precision ID Global-Filer NGS STR Panel, 392
Precision ID identity panel, 79–80, 487–488
Precision ID Identity/Ancestry Panels, 392
Precision ID mtDNA Control Region Panel, 79, 487–488
Precision ID mtDNA Whole Genome Panel, 79, 392, 503–504
Precision ID mtDNA Whole Genome Panel v.2.2, 503–504
Precision ID system coupled with the Ion Torrent MPS, 339
Primary transfer of biological evidence, 23
Principal component analysis (PCA), 134–139, 153–154
  analyses of 17 continental populations based on 40 AI-SNPs, 160–161
Principal components (PC), 153–154
Probabilistic genotyping software (PGS), 478–479
Probability of exclusion (PE), 130, 220
Probability of identity (PI), 182–183
Proline-rich protein HaeIII subfamily 2 protein, 434
Promega PowerPlex 46GY System, 181
Promega PowerSeq 46GY System, 89
Promega PowerSeq CRM Nested System, 491
Promega PowerSeq Mito System, 400
Promega's PowerSeq system, 392
Promissory entities, 560–561
Proof-of-principle study, 350
Protamine-1 (PRM1), 431–432
Protamine-2 (PRM2), 431–432
Protamines, 431–432
Protein biomarkers, 433
Protein Kinase (ProK), 63
Protein markers, 434–435
*PRSS*53 gene, 319
Pyrosequencing, 5–6, 38, 176, 340, 348–349
  based age determination assays, 347
  detecting DNA methylation with, 189–190

Q

Q20+ sequencing kit, 94–95
Qiagen Microbial DNA quantitative PCR Assay Kit, 381
Quality assurance standards (QAS), 490
  Standard 8.2, 491–492
Quantitation methods for microbial DNA, 381
Quantitative data, 254–257
Quantitative polymerase chain reaction (qPCR), 381
Qubit instruments, 381

R

Raman spectroscopy coupled with chemometrics, 425–426
Random Forest, 164
Random match probability (RMP), 105–106, 129
Raw read accuracy, 94–95
Raw sequence data, 6–7
Read counts, 516–517
Real-time polymerase chain reaction (rtPCR), 64–65, 413–414, 425–426
Relative fluorescence units (RFUs), 495–496
Reproducibility test, 177–178
Restriction fragment length polymorphisms (RFLP), 28, 243
Revised Cambridge Reference Sequence (rCRS), 204, 501
Ribosomal RNA (rRNA), 482
RNA analysis, 46
RNA profiling, 188
RNA sequencing, 388, 417
RNA testing, 357–358
RNase-free tubes, 61
Roche 454 systems, 531
Roche Diagnostics' GS Junior System, 183
Roche GS FLX Titanium, 74–75
Rooted trees, 154–155
Routine forensic DNA analysis, 258–259
RSID-Semen lateral flow assay, 413

S

SALIgAE reagent, 413, 435–436
Saliva, 430–431, 433–434
Salivary RNA, 434
Sanger sequencing, 35–36, 76, 290–292, 300, 471–473
  next-generation sequencing (NGS) *vs.*, 59
  technology, 73
Scaled-up GridION, 471
Scanning electron microscopy (SEM), 430–431
SCGN gene, 348

Scientific Working Group on DNA Analysis Methods (SWGDAM), 413, 486, 490–491, 500–501, 517
  interpretation guidelines for autosomal STR typing, 495–500, 498t
  interpretation guidelines for mitochondrial DNA analysis, 500–502
  validation guidelines, 493–495
    developmental validation process, 493–494
Second-generation NGS systems, 74–75
Second-generation sequencing, 37–39, 253, 388, 473
Secondary transfer of biological evidence, 23–24
Semen-specific markers, 431–432
Semenogelin (Sg), 435
Semenogelin-1 (SEMG1), 431–432
Semiconductor sequencing, 74
Sepsis, 368
Sequence-based polymorphism, 243–244
Sequencing by hybridization (SBH), 37–38
Sequencing by synthesis (SBS), 37–38, 85, 225, 245–253, 418
Sex chromosomes, genetic markers located on, 125–126
Sexual assault nurse examiner (SANE), 21–22
Sexual assaults (SA), 22, 30, 452
Sexual chromosomes, 35
Shedder status, effects on, 26–31
Short tandem repeats (STR), 3, 26–27, 35, 59–60, 73, 85, 101–103, 106, 125–126, 133–134, 175, 185, 220, 280, 289–290, 311, 357, 387, 473–474, 485, 491, 532, 539, 551
  analysis, 41–43, 99–100
  based tool, 312–313
  direct and indirect transfer of biological evidence, 23–26
  effects on transfer and shedder status, 26–31
  genotyping, 88–89
  markers, 222–223, 243–244
  new awareness, 32–33
  next-generation sequencing advantages for, 254–258
  profiling, 311
  sequencing, 10–11
  touch DNA, 22–23
    and wrongful convictions, 31–32
Shotgun method, 74
Signature microflora, 436
Silica method, 201–202
Single base extension (SBE), 315
Single molecule sequencing (SMS), 473
Single nucleotide polymorphisms (SNPs), 3, 37–38, 59, 88, 101–103, 106, 134–140, 186–187, 243–244, 283, 312, 357, 387, 451, 476, 485, 532, 539, 553
  analysis, 41–43
  calling, 109
  characterization and origin of, 244
  conventional single nucleotide polymorphism genotyping assays, 245
  forensic application, 258–267
  genotyping, 89–90
  limitations, 267–268
  markers, 106–107, 223
  next-generation sequencing (NGS)
    advantages, 254–258
    conventional genotyping assays, 245–258
  typing, 289–290, 397–398
Single-molecule real-time (SMRT) sequencing, 40
Single-nucleotide addition (SNA), 37–38
Single-site identical by state baiting, 524
Skin cells, 22–23
Skin color, 297–298, 313
Skin prediction, 315
SLC24A5 gene, 313–314
SLC45A2 gene, 313–314
Small proline-rich protein 3 (SPRR3), 433
Smallest device, 87–88
SNaPshot, 348–349, 474–475
  approach-based panel, 259–260
  FDP System, 299
Snipper tools, 315
Soil microbial DNA in forensic science
  challenges, 383
  extraction methods for obtaining microbial DNA from soil and water, 377–381
  geolocation in United States, 382–383
  next-generation sequencing (NGS)
    microbial DNA analysis, 381–382
    microbial profiling of soil and water, 375–377
Solexa Genome Analyzer, 471
SOLiD R, 473
SOLiD second-generation sequencing system, 75
Solid-phase bridge amplification, 37
Solid-phase template walking, 37
Standard operating procedures (SOP), 377, 490
Staphylococcus aureus, 368

State-regulated DNA databases, 554
Statistical analytical methods of population genetic study, 151–157
  F test, 152–153
  genetic differentiation index (F_{ST}), 151–152
  multidimensional scaling, 153
  phylogenetic tree, 154–155
  principal component analysis, 153–154
  structure analysis, 155–157
Stillwater, 32
Streptococcus salivarius, 436
Streptococcus sanguinis, 436
Substrate material, effects of, 24–25
Supported Oligonucleotide Ligation and Detection System 2.0 platform, 5–6
Surveillance capitalism, 560
Sweat, 431, 435

T

Tamm-Horsfall glycoprotein (THP), 431
Targeted (re) sequencing, 74
Targeted genomic resequencing, 8–9
Taxonomy-independent deep neural network approach, 436–437
TCHH gene, 319
Technology evolution, timeline of, 471–473
Teichman crystal test, 428–429
Template preparation, 6
Ten-eleven translocation (TET) family, 321
Tertiary transfer, 24
  of biological evidence, 24
Thermo Fisher Scientific Precision ID mtDNA Whole Genome Panel, 491
ThermoFisher's Ion Torrent PGM, 475
Third generation sequencing, 39–40, 325, 473
  platforms, 253
  systems, 75–76, 85
Time of colonization (TOC), 481
Time since death (TSD), 362–363
  determination, 362–364
Tissue-specific differentially methylated regions (tDMRs), 337
Total internal reflection fluorescence (TIRF) microscopy, 38–39
Touch DNA, 22–23, 31–32
Towson University (TU), 546–547
Trace DNA, 22–23
Traditional capillary electrophoresis-based assays, NGS *vs.*, 254
Traditional DNA profiling, 311
Traditional privacy models, 556
Transcript-based approaches, 419–420
Transcript stable regions concept (StaRs), 414–417
Transfer status, effects on, 26–31
Transfer type, 24
Transglutaminase 4 (TGM4), 431–432
*TRIM*59 gene, 324
True SNPs, 244
Tuberculosis (TB), 197
23andMe, 515, 543–544
Twin differentially methylated CpG sites (tDMSs), 191
TYR gene, 313–314
*TYRP*1 gene, 313–314

U

Ultraviolet (UV), 392
  irradiation method, 201–202
  light exposure, 293–294
  radiation, 317
  UV-Vis spectroscopy, 64–65
Ultraviolet C (UVC), 198
Union-exon-based approaches, 419–420
Uniparental disomy (UPD), 230
Unique molecular identifiers (UMIs), 94–95
United Kingdom (UK), 489
United States (US), 514–515
  application of microbial DNA analysis in geolocation in, 382–383
Unrooted trees, 154–155
Urine, 431, 434–435
Uromodulin, 434–435
US Combined DNA Index System (CODIS), 29, 181, 261–262, 286, 489, 539–540
US Department of Agriculture-National Animal Health Laboratory Network, 48
US Department of Justice, 505, 542
Use of NGS in forensic sciences, 76–81

V

Vaginal fluid, 430, 433
Validation of next-generation sequencing-based DNA analysis methods, 490–505
  developmental validation process, 494
  Federal Bureau of Investigation quality assurance standards, 491–493
  national DNA index system (NDIS), 491

Validation of next-generation sequencing-based DNA analysis methods (*Continued*)
next-generation sequencing (NGS)
peer-reviewed publications on validation studies, 502–505
scientific working group for DNA analysis methods
guidelines, 493–495
interpretation guidelines, 500–502
δ value, 132
Variable number of tandem repeats (VNTRs), 289
Verogen ForenSeq DNA Signature Prep Kit, 491
Verogen ForenSeq MainstAY SE kit, 181
VISible Attributes through GEnomics (VISAGE), 44, 262–263, 324, 350, 478
initiative and future of forensic DNA phenotyping, 324–325
VisiGen Biotechnology approach, 40

W

Water, extraction methods for obtaining microbial DNA from, 377–381
Water microbial DNA in forensic science
challenges, 383
extraction methods for obtaining microbial DNA from soil and water, 377–381
geolocation in United States, 382–383
next-generation sequencing (NGS)
microbial DNA analysis, 381–382
microbial profiling of soil and water, 375–377
quantitation methods for microbial DNA, 381
WeGene, 543–544
Whole exome sequencing (WES), 266, 476
Whole genome amplification, 376–377
Whole genome sequencing (WGS), 59, 263, 539–540
datasets, 557
WNT10A gene, 319
Wound age estimation using miRNA, 364–366, 366t
Wright–Fisher model, 122
Wrongful convictions, 31–32

X

X chromosome SNP/InDel (X-STR/InDel), 126–127
X chromosome STR (X-STR), 126–127, 487
X-chromosome genetic markers, 126
X-ray diffraction (XRD), 375

Y

Y chromosomal STRs (Y-STRs), 77, 126–127
profiling, 357
Y chromosome, 43
specific region, 126
Yersinia pestis, 92, 376
Yoogene, 543–544

Z

Zinc finger proteins, 190–191

Next Generation Sequencing (NGS) Technology in DNA Analysis

Edited by Hirak Ranjan Dash, Kelly M. Elkins and Noora Rashid Al-Snan

Next Generation Sequencing (NGS) Technology in DNA Analysis explains and summarizes next generation sequencing (NGS) technological applications in the field of forensic DNA analysis. The book covers the transition from capillary electrophoresis (CE)-based technique to NGS platforms and the fundamentals of NGS technologies, applications, and advances. Sections provide an overview of NGS technology and forensic science, including information on processing biological samples for forensic analysis, sequence analysis, and data analysis software as well as the analysis of NGS data. The book explores the valuable applications of NGS-based forensic DNA analysis and covers the validations and interpretation guidelines of NGS workflows.

With chapter contributions from an international array of experts and the inclusion of practical case studies, this book is a useful reference for academicians and researchers in genetics, biotechnology, bioinformatics, biology, and medicine as well as forensic DNA scientists and practitioners who aim to learn, use, apply, and validate NGS-based technologies.

Key Features

- Describes the fundamentals of NGS and its advances for forensic applications
- Explains the transition from CE-based technique to NGS technology
- Includes case studies related to NGS and DNA fingerprinting
- Explores the future use and applications of NGS technologies

About the Editors

Dr. Hirak Ranjan Dash is an assistant professor of forensic biotechnology at National Forensic Sciences University, Delhi Campus, India. He obtained his PhD in Life Science from the National Institute of Technology, Rourkela. Before joining academia, he served as a DNA expert at Forensic Science Laboratory, Madhya Pradesh, India. He has had more than 12 years of research experience. His research interests include forensic DNA analysis, genetic markers, forensic microbiology, and NGS. He has published 50 research papers and 10 books to date. He is the recipient of a research fellowship from the Indian Academy of Science. He is a pioneer in India in working on NGS-based forensic DNA analysis. He is a life member of the International Society of Forensic Geneticists and the Association of Microbiologists of India. He was featured in the list of the world's top 2% researchers in 2021.

Dr. Kelly M. Elkins is a Professor of Chemistry at Towson University where she teaches in the Masters of Science in Forensic Science Program and is the co-Director of the TU Human Remains Identification Laboratory. Previously, she was the Director of Forensic Science at Metropolitan State College of Denver. She has her Bachelor of Science in Biology, Bachelor of Arts in Chemistry, and her PhD in Chemistry with an emphasis on Biochemistry. Focused on the Forensic DNA field, she has developed a network of contacts in crime laboratories throughout the U.S. and internationally. She is a Fellow of the American Academy of Forensic Sciences and a member of the American Chemical Society. She has published almost 70 books, book chapters, and journal articles. She consults on forensic cases.

Dr. Noora Rashid Al-Snan is the head of the Biology and DNA Forensic Laboratory for the Directorate of Forensic Science, Ministry of Interior, Kingdom of Bahrain. She has a BSc in molecular biology–biochemistry from Kuwait University and an MSc in medical biotechnology from Arabian Gulf University, and a PhD in molecular medicine from Arabian Gulf University, specializing in medical genetics. She has authored publications on forensic genetics and genetic genealogy studies. She has published several case work studies on terrorism.

ISBN 978-0-323-99144-5

9 780323 991445

ELSEVIER

ACADEMIC PRESS

An imprint of Elsevier

elsevier.com/books-and-journals